CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

21-276

STATISTICAL REVIEW(S)

Statistical Review and Evaluation

NDA/ Drug Class:	21-276 / 6S
Name of Drug:	Estrostep® (norethindrone/acetate/ethynyl estradiol) Tablets
Applicant:	Parke-Davis Pharmaceutical Research
Indications:	Treatment of moderate acne vulgaris in females age 14 or greater who desire contraception.
Documents Reviewed:	Volumes in FDA Electronic Data Room, including SAS data sets, plus paper volumes 1.2, 1.33-155, submitted July 3, 2000.
Medical Officer:	Joseph Porres, M.D., Ph.D. (HFD-540)
Statistical Reviewer:	Steve Thomson (HFD-725)

Contents:

I. Introduction and Background

The sponsor proposes to use Estrostep® (norethindrone/acetate/ethynyl estradiol) Tablets, indicated for the prevention of pregnancy, in the treatment of moderate acne vulgaris in [redacted] to 49-year-old women.

The sponsor claims that combined oral contraceptives are the most effective and widely-used reversible method of birth control. Estrostep is a combination oral contraceptive that contains 1 mg norethindrone acetate (NA) and a graduated dose sequence of ethinyl estradiol (EE) over a 21-day period. Estrostep was approved by the FDA for prevention of pregnancy on October 9, 1996,

According to the sponsor: "Nearly all of the currently available acne treatments target the lesions that form as a result of increased sebum production. In contrast, oral contraceptives lower the androgen levels that trigger increased sebum production. While there are other pharmacologic means of decreasing androgen secretion or action, they have dose-limiting systemic side effects and are not approved for the treatment of acne. Oral contraceptives reduce androgen-mediated sebum production by 2 mechanisms. They suppress gonadotropin secretion, decreasing ovarian androgen production, and they increase the concentration of SHBG, reducing the amount of free testosterone available."

II. Experimental Designs

To support the claim of efficacy and safety the sponsor provided data from two nearly identical studies (labeled 376-403 and 376-404, respectively). Both studies were 6-month, randomized, double-blind, parallel group, multicenter studies assessing the efficacy and safety of Estrostep versus vehicle in the treatment of moderate acne vulgaris. In Study 376-403, 298 subjects were enrolled, of whom 188 were considered to have completed the study. In Study 376-404, 295 subjects were enrolled, of whom 215 were considered to have completed the study.

From the within center runs in allocation to treatment group, it is apparent that patients were randomized in blocks of four patients. Simple counts of run patterns showed no particular pattern that suggested problems with the randomization to treatment.

Measurements were tied to the subject's menstrual cycle, with a treatment period consisting of six, 4-week cycles. According to the protocol the subjects were to be seen for randomization on menstrual cycle day 1 through 6, and then once during cycle days 17 to 24 for cycles 1 and 3-5, and cycle days 18 to 21 for cycle 2, defining the windows for visits 3-7. For cycle 6, three visits were scheduled at cycle days 1 to 7, 8 to 14, and 18 to 21 (visits 8-10). Two particular statistical problems with the designs as implemented were the large number of subjects in one center in Study 376-403, and the fact that in both studies a large proportion of centers had relatively few subjects.

II.A. Endpoints

Primary Efficacy Parameters:

· Total number of acne lesions, inflammatory lesions, and comedones from baseline to study exit.
· Facial Acne Global Assessment at study exit.

A prima facie case for a claim of efficacy is established when the sponsor achieves statistically significant differences in favor of treatment in two of the three lesion counts, AND in the dichotomized Facial Global Assessment (see below for details on the dichotomization).

The original protocol specified that the lesion counts were to be evaluated as change (from baseline) scores, i.e. the difference of the observed measure from baseline. That is a reasonable measure, but in the sponsor's analysis it was the only primary measurement. In the analysis presented in this review equal weight will be given to the percent change from baseline (i.e., 100 times the change score/ baseline score). This percent change would seem to give adequate information about the change score. Under such circumstances it was felt that lesion counts would be more appropriately evaluated as actual counts, rather than as the change scores proposed in the protocol. Statistically, provided the baseline covariate lesion count is included in the model, all tests of differences among treatment, sites, or interactions would be identical to those provided by the change score model. That is, except that the covariate slope is decremented by one unit in the change score model, the models are equivalent. Hence this was not considered a major change to the protocol.

To reiterate, the lesion counts were evaluated both as an actual count of lesions and as percent improvement from baseline. The latter was computed as the (baseline - exit score)/baseline in terms of percents. The Facial Acne Global Assessments were each evaluated on a seven point scale:

1 = Absent	5 = Moderate
2 = Minimal	6 = Marked
3 = Mild	7 = Severe
4 = Mild to Moderate	

According to the sponsor's protocol these global assessments were to be dichotomized into a binary endpoint defining "improved" (i.e., "success"), as a score of absent, minimal, or mild (1-3), versus "not improved" (i.e., "failure"). Note that the parenthetical labels will be used in this report, since the sponsor's labels seem to suggest a response relative to baseline, whereas the Division of Dermatological and Dental Drug Products prefers an assessment relative to the physicians general experience, disregarding any possible baseline evaluation. In fact, in this study the first evaluation of the Facial Acne Global Assessments was not performed until visit 3, after initiation of treatment, and hence can not reflect a change relative to baseline.

II.B. Sponsor's Analysis/Pooling Centers

The protocols for each of the studies proposed that the original lesion counts would be analyzed by an analysis of covariance (ANCOVA) on the (change) scores, with factors for treatment group and investigator. However, as noted above, the actual score, not the change score will be used here. Baseline score was to be used as a covariate. The protocol proposed treating percent change from baseline as a secondary endpoint, analyzed with a similar ANCOVA. Further, the protocol specified that the investigator global assessment was to be evaluated using a Cochrane-Mantel-Haenszel (CMH) test stratified on investigator (i.e., site). As noted above, responses of "absent", "minimal", and "mild" were to be considered as a treatment "success", other responses as a treatment "failure".

Table 1 below shows the baseline subject counts for the centers in both studies. One problem with the proposed analyses is the badly unbalanced distribution of sites/centers. The CMH statistic uses a hypergeometric probability model for the distribution of responses within each stratum. When a row or column has a zero marginal there is no variance and the stratum is dropped from the computation of the CMH statistic. So the large number of small centers may be problematical for the CMH statistics. Unbalanced data may also cause problems for ANOVA/ANCOVA analyses. For example, for balanced data, with equal numbers of observations in each cell in the design, ANOVA methods are generally distributionally quite robust. But when there is severe unbalance in center sizes the tests of effects become ever more sensitive to deviations from normality in distribution. Further, for analyzing fixed-effects ANOVA/ANCOVA models there are various approaches to defining the spanning subspaces associated with a labeled effect in the design. For balanced data the usually proposed methods all will give exactly the same results. But when the data are badly unbalanced as here, these approaches can give very disparate results.

The sponsor's analysis of the ANCOVA models described above was based on using so called Type III sums of squares, where all centers/investigators are weighted equally. Thus if centers are disparate in size, then individual patients can be weighted quite unequally in the analysis. The sponsor also proposed an analysis of the Facial Acne Global Assessments using the CMH statistic stratified on center, with similar problems as noted above. To enable one to use Type III sums of squares in ANOVA, and to use the corresponding CMH statistic as proposed in the protocol, it would seem to be important that center effects be defined so that centers are not so disparate in size.

This argument above suggests that we pool the smaller centers into larger "Pooled" centers. Arbitrarily it was decided that centers with fewer than 15 subjects would be considered "small" and hence pooled. The pooling of these smaller centers could be done a number of ways, say by demographics or on the basis of other knowledge. A reasonable alternative, which was used here, would be to pool them on the basis of the baseline values of the primary endpoints. It could be hoped that this would group centers with similar subjects and similar evaluations. Note that the physician's global facial acne assessment was not evaluated until after initiation of treatment, and hence was not considered an appropriate measure on which to pool the centers. Thus it was decided to group centers on the basis of the two functionally independent lesion counts, i.e. to group centers on the basis of a "cluster analysis".

Hierarchical or agglomerative cluster analysis forms clusters of points in some space on the basis of the distance between points. An initial cluster is formed by taking the two closest points. Then the next

two closest points are joined, one of which could be the initial cluster. At each step, points or clusters are grouped together (hence the label "agglomerative"). The clustering stops when all points are pooled together (hence the label "hierarchical"). Various hierarchical clustering procedures are defined by different methods of defining distance between points. The successive grouping of points into clusters is usually represented as a tree-like dendrogram, showing the grouping of points and the relative distance at which points were pooled. However, SAS® does not produce dendrograms directly, only a diagram equivalent to a dendrogram that is messy and is not reproduced in this report. Much of the same information in the dendrogram can be gleaned from the simple table of formed clusters.

Text Table 1 below displays the number of subjects at baseline (here last visit in the nominal visits 1 or 2) and the corresponding within center mean inflammatory lesion count and comedone count for both studies. The numeric center labels were those assigned by the sponsor. The "pooled" centers are those used henceforth in the analysis. Note a typical detailed sequence of formed clusters in both studies using one of the cluster methods is presented in Table A.1 of the Appendix.

Table 1. Number of Subjects per center and Mean Lesion Counts at Baseline

Study 376-403				Study 376-404			
Center	Inflammatory Lesions N	Inflammatory Lesions Mean	Comedones Mean	Center	Inflammatory Lesions N	Inflammatory Lesions Mean	Comedones Mean
006	15	26.6	45.3	001	20	27.2	49.6
009	131	29.0	41.2	003	35	26.1	30.7
010	32	31.6	59.6	005	17	24.9	30.8
012	16	30.9	58.6	007	29	36.7	61.5
014	31	35.3	40.9	010	49	31.3	34.0
				011	24	25.6	30.8
				012	31	26.7	41.6
				014	26	30.8	45.8
Pooled 1	29	28.9	67.5	Pooled 1	29	27.8	31.2
001	3	20.7	66.0	002	6	29.0	30.5
002	7	25.4	58.0	006	8	30.5	35.9
005	3	25.3	58.7	009	13	26.8	30.2
015	8	37.9	92.9	019	2	20.0	21.5
017	6	27.3	54.3				
Pooled 2	43	26.4	40.0	Pooled 2	33	32.0	51.6
004	7	26.3	37.4	004	9	32.4	50.7
007	10	24.8	48.1	013	8	33.4	46.9
008	8	23.6	41.1	015	3	22.3	48.7
013	7	27.6	29.7	016	7	34.6	63.1
016	6	28.0	44.0	017	4	28.3	45.0
018	5	30.6	35.6	018	2	37.5	51.5

Originally it was hoped that three roughly equally sized, pooled centers could be created for each study. However, it seemed clear from the dendrograms (not displayed in this report) from several different hierarchical/agglomerative clustering procedures that in both studies the data naturally form two reasonably balanced centers. Actually, in Study 376-403, center 015 is pooled last, indicating that it's baseline means are disparate from the others that were agglomerated in that study. However, for convenience it was placed in pooled center 1.

One way of assessing the effect of this pooling would be to investigate its effect on within center by treatment error variances. Under the normal model, if one ignores the baseline covariate, these should be independent, with a chi-square distribution. We can estimate the overall variance from the pooled within center variances s^2_{pool}. Then we can plot the within center by treatment sum of squares divided by s^2_{pool}. The expected value of this ratio should be approximately n-1, the degrees of freedom associated with the sum of squares. Since the variance of a chi-square random variable is twice its degrees of freedom, a plot of this ratio versus its degrees of freedom should show a fan-shaped pattern. The primary goal of this plot was to see if the values associated with pooled centers were disparate from those associated with other centers. Typical, example plots appear as Figures 1 and 2 in the Appendix. Note that these plots do indicate some overdispersion (i.e. more points outside the standard error limits than would be expected from strict normality), but the values associated with pooled centers seem to be quite consistent with those from other centers.

Since the pooling of centers using the cluster analysis described above was a post hoc adjustment of the data in the experiment, it was of interest to investigate the sensitivity of the results to the specific pooling being used. That is, it is of some interest to compare the results from the pooling used above to other possible groupings of centers. Table A.2 in the Appendix illustrates the effect on the significance level of the test of treatment differences of different ways of randomly pooling those centers with less than 15 subjects into two "super-centers." Though only 20 examples are shown there, several hundred simulations were run. It is apparent that the impact of the clustering has little impact upon the final results, and apparently a random pooling would have performed just as well as that derived from the cluster analyses.

One approach recommended by the Division of Biometrics for the analysis of extremely unbalanced data, as in these studies, was a mixed model approach to testing treatment effects. These were applied to models without pooling centers, rather treating sites as a random effect. An important assumption in such an analysis is that dropouts and missing observations are approximately missing at random (MAR),), i.e. the probability of being missing may depend upon factors and covariates including time, but can not depend upon the actual value of the response at that time point. Under this MAR assumption the mixed model can be used to adequately model response over time even with the relatively high drop out rate observed here.

In doing such a mixed model analysis, first one must assume a covariance structure for the responses. Typically this reviewer would first test effects assuming an exchangeable model, as was done here, and then simplify, and finally verify the exchangeable model assuming other various covariance structures. The results from the initial run assuming exchangability (equal correlation) are exemplified in Appendix Tables A.3 and A.4. Only the output for the actual lesion counts is given there. Results for the percent change in respective lesion count were similar. Further, the results associated with subsequent models, with reduced fixed effects and alternative covariance structures, were generally consistent with those presented in these tables. Note that these results, as well as those for mixed models stratified on time, are, in turn, generally consistent with the results of the fixed effects models using the pooled centers, and hence are not discussed further.

II.C. Statistical Methodology

In an ANOVA or ANCOVA analysis, when there are no empty cells in the design and when the interaction term is included in the model, the test for main effects using Type III sums of squares is a simple comparison of marginal adjusted cell means. To this reviewer such a comparison is quite interpretable even if the interaction is statistically significant. However, when the test for treatment by center interaction is significant and large, the generalizability of the results to other centers is questionable. Still, when interest is restricted to the specified centers or when the interaction tests are not particularly large relative to main effects, as here, it is this reviewer's opinion that the Type III tests of differences seem particularly appropriate. When no interaction term appears in the model, the Type III contrasts in cell means are orthogonalized to the corresponding interaction contrasts, and seem somewhat less easily interpretable. Thus, when using Type III sums of squares, this reviewer would normally recommend including the interaction term in the model, whether interaction is present or not.

However, the protocol specified that lesion counts would be analyzed with an analysis of covariance with baseline lesion count as a covariate and center and treatment as factors. Apparently interaction was only to be included as a factor in secondary analyses. This reviewer generally believes that the protocol should be followed unless there is some fairly strong reason not to do so. Thus when the test for interaction was statistically non-significant, the interaction was NOT included in the model. The specific rule used here was that if the significance level for the test of interaction was highly non-significant (here $p \geq 0.25$), the model without interaction as specified in the protocol was used. When the significance level of the test for interaction had $p < 0.25$, the interaction term was retained in the model, and the effect of interaction was addressed. In Study 376-404 several interactions have significance levels less than this 0.25 level. Note that if treatment effects were zero, we would expect an about half the centers to have a comparison between treatments where the treatment group is somewhat superior to vehicle and about half where the vehicle is superior to the treatment group. When one treatment is strongly superior to another we would expect that the same superiority of treatment over vehicle would hold within each center. However, reversals, i.e. qualitative interactions, can and do occur by chance. The probability of such reversals increases when the treatment effect is relatively small, as is true here. In Study 376-404 there are several such apparent reversals (see the discussion in that section).

Note that in the labelling the sponsor was making a claim of early relief. To investigate this, similar tests were made at each visit, both using observed data and LOCF imputed data. To improve comparability of the models over the time points, the same model, with interaction, was used for each of these tests (see Appendix Tables A.9-A.11 and A.14-A.16), even though at most time points and endpoints the test for interaction was statistically nonsignificant. Although the p-value unadjusted for multiple comparisons is presented in the table, to control the Type I error associated with analyzing multiple visits for each lesion count or global assessment these should be cast as a so-called "closed" family of tests. That is, the results at week 9 are evaluated if and only if the test of difference between vehicle and Estrostep are statistically significant at week 10. Similarly the week 8 results are evaluated if and only if the test of difference between vehicle and Estrostep are statistically significant at week 9. If at any visit the differences are not statistically significant, the analysis stops and reports any observed statistical significance up to that time point. Note that there are multiple endpoints, but because of the dependence of the total lesion count on the inflammatory and noninflammatory lesion counts, and since we require statistical significance on two of the three lesion counts, no correction for the multiple response variables is needed.

One assumption of the mixed model analysis cited above was the assumption that data were MAR. To address this directly requires the response at the time the subject drops out. But the difficulty is that the actual value is not observed for those who drop out. A way to address this issue was to use the prior value of the lesion count as a proxy for the current value. When this was done in a logistic model at each visit, though the results are not displayed here, the probability that a patient among those remaining at that visit would drop out, modelled as a function of site, treatment, baseline, and the lagged total lesion count, showed that the effect of the lagged prior count was always statistically quite nonsignificant in Study 376-403. Again, while the results are not displayed in this report, it was found that the effect of the lagged value was only barely statistically significant once in Study 376-404. Since this is the sort of artifact we would expect when a number of tests are run, it does appear that the assumption of no effect of the lagged value is generally accepted. It is true that accepting the null hypothesis does not show that the data are MAR, but consistent failure to reject is quite consistent with the assumption that the data are MAR.

Further, to investigate if treatment had any effect on whether or not a subject was an early dropout, a logistic model for whether or not a subject was a dropout was fit. Factors were baseline total lesion count, treatment, and site. Again, though results are not shown here, neither study showed any differences due to treatment that were even close to being statistically significant (though there were statistically significant center/site differences).

II.D. Assessing Validity

The Medical Officer expressed an interest in validating the investigator global evaluation, and proposed a table showing a count of those subjects who had at least a one unit improvement (i.e., a decrease) in the global evaluation from the first measurement (visit 3). Similarly the subset of subjects with at least a two unit improvement, and the subset of those with at least a three unit improvement in their global evaluation were also tabulated. See Tables A.5 and A.6 in the Appendix for these tabulations. In addition the corresponding mean decrease in total lesion count is displayed for each of these decreasing subsets of subjects. One use of this table is to show that as the global evaluation improves there is generally a corresponding decrease increase in total lesion count.

A more formal approach would be to assess the validity the global evaluation using a criterion referenced measure of validity. As seems usual in such a formulation we will use the association of the global evaluation scores with variables having high "face" validity, i.e., lesion counts. Again, as is typical, it seems natural to use the within subject correlation between the global evaluation and the various lesion counts as measures of validity. However, even with normally distributed data, the distribution of the correlation coefficient is skewed, and quite non-normal. A first order correction for this non-normality is to transform the computed correlation r using the so-called Fisher z-transform: $0.5 \log((1+r)/(1-r))$. Asymptotically the transformed values are distributed as approximately normal random variables with variance $1/(n-3)$, where n is the number of observation pairs used to compute the correlation. So it makes sense to limit the computed correlations to subjects who had four or more observed cases.

If subjects could be assumed to be independent, a weighted sum of the z-transformed values would be the appropriate estimate of the transformed population correlation. However subject correlations can be expected to be associated within site. Because of this within site correlation, the z-transformed observed correlations were analyzed using a weighted mixed model with site and subject variances, with each observed correlation having the appropriate weight n-3. After fitting to give the estimated mean correlations, these point estimates of the population mean and the corresponding 95% confidence intervals were back transformed to the original correlation scale to give:

Estimated Correlations of Lesion Counts With the Physicians' Global Evaluation

	Study 376-403		Study 376-404	
Lesion Count	Estimate	Confidence Interval	Estimate	Confidence Interval
Inflammatory Lesions	0.61	(0.52 , 0.69)	0.61	(0.52 , 0.70)
Comedones	0.43	(0.27 , 0.57)	0.44	(0.31 , 0.56)
Total Lesions	0.59	(0.48 , 0.69)	0.61	(0.51 , 0.70)

Note that these estimated correlations are quite consistent across studies. The general rule of thumb in most evaluation studies seems to be that a correlation above 0.5 is considered "large". Thus, it seems apparent that the investigator's global evaluation is associated with the lesion count, particularly inflammatory lesions, and presumably could be considered a valid measure. However, it also worth noting that while a 0.6 correlation may be considered as a high value in evaluation studies, that does mean that roughly only 36% of the variation in the investigator's global evaluation is "explained" by the inflammatory lesion count.

III. Primary Efficacy Results:

III.A. Protocol 403

All patients were female with demographic characteristics as summarized in Table A.7 of the Appendix. As noted in section II. B. above, neither study seems to show a differential drop-out rate due to treatment. Also, because the centers above were pooled in this reviewer's analysis, and thus treated differently than in the sponsors analysis, no attempt to compare the sponsor's analysis and this reviewer's analysis will be made in this report. However, despite the warnings of section II.B. above, the sponsor's results were generally quite consistent with this reviewer's analysis

The Type III sums of squares used in this analysis essentially give equal weight to centers, and hence down weight the effect of the unusually large center 009 (see Text Table 1 above). Since there were apparent treatment differences in center 009 this will tend to be conservative. Note as explained in section II above, visit 10 was nominally the final visit, conducted during days 18-21 of the sixth month menstrual cycle. Those subjects tabulated under that visit are the group of subject "completers," who completed the study. The point "End of Study" ("eos") is defined for the more inclusive set of ITT subjects as this final visit 10, or the subjects last valid visit prior to this visit 10, where the last measurement is imputed forward using "last observation carried forward" technology.

The following table, Text Table 2., displays the mean actual counts and the corresponding least squares means (a.k.a. "population marginal mean") of actual counts of inflammatory lesions, comedones, and total lesions, as well as the per cent change from baseline for each of these measures, both for those who completed to visit 10 and to the end of study for the ITT population (i.e. all subjects dispensed treatment). Least squares means (i.e., "LS means") are estimates of the corresponding cell mean had the data been balanced, with equal numbers of observations per cell. Note that the Type III sums of squares used in this report are based on simple contrasts of these LS means, and hence this reviewer would recommend emphasizing these LS means instead of the sample means as measures of treatment effect. However, LS means treat centers equally, and hence downweight the subjects in the larger centers. Since centers are badly unbalanced, the mean of all the scores may also be relevant, and is provided here. Again, for this end-of-study measure ("eos") missing responses were imputed using last-observation-carried-forward technology. In general, following apparent ICH guidelines, we emphasize the LOCF results in the ITT population. However, it does appear that in this particular case, the care taken to pool centers and perform nominally more appropriate tests had no particular impact on the final results.

Significance levels are given for the corresponding test of treatment differences using both an analysis of covariance model with the baseline score as a covariate and factors for site and treatment. Note that while several measures had relatively large investigator effects, there were no statistically significant treatment by investigator interactions at any of these time points. When these models were run with an interaction term, all interactions were not quite statistically nonsignificant (smallest $p \geq 0.7204$). Although there can be problems in deleting statistically nonsignificant effects, this seems to justify the use of the restricted model without interactions in this study as originally specified in the protocol. Hence, all models in the table below specified the baseline score as a covariate with site and treatment (but not interaction) as factors.

Table 2. Study 403: End of Study Lesion Counts

Includes sample sizes, means and LS means, standard errors for each, and p-values from ANCOVA tests of treatment differences.

		Visit			
		10 (Completers)		eos (ITT)	
		Vehicle	Estrostrep	Vehicle	Estrostrep
N		86	100	148	149
Inflammatory Lesion Count	Mean	16.8	15.4	19.1	16.5
	Std Err	1.2	1.1	1.0	0.9
	LS Mean	14.8	13.7	18.7	16.3
	Std Err	1.2	1.2	1.0	1.0
	p-value*	0.4290		0.0515	
Total Comedones	Mean	30.3	27.0	34.3	30.0
	Std Err	2.4	2.4	2.1	2.0
	LS Mean	29.9	26.2	36.2	30.9
	Std Err	2.1	2.1	1.8	1.8
	p-value*	0.1525		0.0197	
Total Lesion Count	Mean	47.0	42.4	53.4	46.6
	Std Err	3.2	3.1	2.8	2.5
	LS Mean	44.8	39.8	55.0	47.3
	Std Err	2.8	2.8	2.5	2.5
	p-value*	0.1501		0.0118	
% Change in Inflammatory Lesions	Mean	41.2	46.6	33.3	41.3
	Std Err	4.4	3.6	3.5	3.2
	LS Mean	47.8	51.9	33.4	41.8
	Std Err	4.4	4.3	3.6	3.6
	p-value*	0.4409		0.0645	
% Change in Comedones	Mean	27.6	41.5	20.1	34.5
	Std Err	5.5	5.3	4.5	4.1
	LS Mean	30.1	39.4	16.7	30.2
	Std Err	5.3	5.2	4.6	4.6
	p-value*	0.1937		0.0187	
% Change in Total Lesions	Mean	35.5	44.6	27.7	38.1
	Std Err	3.8	3.8	3.3	3.1
	LS Mean	39.8	46.0	25.7	35.9
	Std Err	3.8	3.8	3.4	3.4
	p-value*	0.1845		0.0169	

*p-value is for test of treatment differences in LS Means.

In the ITT population differences at the end of the study between vehicle and Estrostep were statistically significant both for comedones (original score $p \leq 0.0197$ and % change $p \leq 0.0187$) and total lesions (original score $p \leq 0.0118$ and % change $p \leq 0.0169$). The differences for inflammatory lesions were not quite statistically significant (though close) for both the original score and the percent change from baseline. Generally, the differences among the subjects who completed to visit 10 were consistent with, but somewhat smaller than among the LOCF subjects at the end of the study. However this former group was roughly a third smaller, and tests of differences were not statistically significant.

Note that while differences between Estrostep and its vehicle in terms of comedones and total lesions were statistically significant at the end of the study, the magnitude of these effects is fairly small. Whether or not the roughly 10-15% or so difference in percent change is practically significant is a decision requiring the clinical expertise of the Medical Officer.

The following table shows the distribution of the investigator facial acne global evaluation. Note that the protocol specifies that the primary endpoint for this measure is a dichotomization of this variable by those who were mild or better.

Table 3. Study 403: Investigator's Global Evaluation

	Visit							
	10 (Completers)				eos (ITT)			
	Vehicle		Estrostep		Vehicle		Estrostep	
	n	%	n	%	n	%	n	%
Absent	.	.	2	2.0	.	.	2	1.5
Minimal	7	8.1	19	19.0	10	7.4	23	17.0
Mild	25	29.1	24	24.0	32	23.7	35	25.9
Mild to Moderate	33	38.4	38	38.0	51	37.8	48	35.6
Moderate	19	22.1	14	14.0	38	28.2	22	16.3
Marked	2	2.3	3	3.0	6	3.0	5	3.7

	Visit							
	10 (Completers)				eos (ITT)			
	Vehicle		Estrostep		Vehicle		Estrostep	
	n	%	n	%	n	%	n	%
Minimal or better	7	8.1	21	21.0	10	7.2	25	16.2
p-value*	0.019				0.021			
Mild or better	32	37.2	45	45.0	42	30.2	60	40.1
p-value*	0.345				0.123			
Mild/Moderate or worse	54	62.8	55	55.0	93	69.8	35	60.3

*p-value is for test of treatment differences in dichotomized response.

Note the statistical significance of the primary endpoint specified by the protocol is 0.123, i.e., not statistically significant. However, a binary split at minimal or better would be statistically significant (i.e., $p \leq 0.021$), and at both time points the observed proportion of "successes" favors Estrostep over its vehicle. Whether or not these effects are of sufficient magnitude to be clinically important is, again, a decision requiring the clinical judgement of the Medical Officer.

Appendix Tables A.9 - A.11 show the response profiles over time, both for subjects who have data at the specified time point and for those imputed using LOCF. As noted in section II.C, to control the Type I error associated with analyzing multiple visits for each lesion count or global assessment, these are set as a closed family of tests. That is, the results at week 9 are evaluated if and only if the test of difference between vehicle and Estrostep are statistically significant at week 10. Similarly the week 8 results are evaluated if and only if the test of difference between vehicle and Estrostep are statistically significant at week 9. If at any visit the differences are not statistically significant, the analysis stops and reports any observed statistical significance up to that time point.

Subgroup Analyses

Of course all patients were female. For lesion counts broken down by age and race groups see Appendix Table A.12. These studies are not powered to analyze data in subsets, and hence only simple descriptive statistics are provided in this table. We can see that for both age groups, the counts and the per cent change scores in comedones and total lesions show some superiority of Estrostep over vehicle. For inflammatory lesions results are somewhat more problematical, but since there was no statistically significant treatment difference in the entire sample, the apparent lack of differences in the subsamples is not surprising. On the other hand, it seems apparent that at least in this study, efficacy is largely limited to whites.

Table A.13 in the Appendix shows a similar breakdown for the global evaluation of acne. Note that for both age groups Estrostep is generally uniformly better than its vehicle, but while Estrostep is clearly superior to its vehicle among white patients, results are somewhat more problematical for Black and Other (mainly Hispanic) subjects.

III.B. Protocol 404

Again, of course, all patients were female, with demographics summarized in Appendix Table A.8.

For the analysis of lesion counts the protocol specified a model with the baseline score as a covariate and factors for site and treatment. However, prior to using this model it is useful, if not mandatory, to check for interaction. In this study the following tests for interaction between site and treatment suggested the possibility of large effects:

P-values of tests for interaction:

	Visit 10 Completers	Eos
Inflammatory Lesions	0.0472	0.2285
Total Lesions	0.1885	0.2392
Per Cent Change in Inflammatory Lesions	0.0415	0.1077

Although details are not presented here, for inflammatory lesions among subjects completing the study to visit 10, and for total lesions among these completers at visit 10 and at the end of the study ("eos"), the interaction was clearly quantitative, i.e. Estrostep was generally equal to or superior to its vehicle, but the degree of superiority varied considerably across centers. In all cases, the effect size for this interaction was considerably less than the test for differences in treatment. For inflammatory lesions at the end of the study, the mean differences were larger, but so was the amount of variation, so that the actual effect size is less than at visit 10. Similarly the interactions for percent change from baseline in inflammatory lesions seem to be quantitative. When there are a number of centers, and treatment effects are fairly small, we would expect some quantitative interaction, perhaps even some qualitative interaction. So is its presence here is tentatively ascribed to an artifact of the experiment. For those measures listed in the table above, the tests for main effects in the following table, Text Table 4, come from models including a term for interaction.

The next table, Text Table 4., displays the simple mean and corresponding least squares mean for the counts of inflammatory lesions, comedones, and total lesions, as well as the per cent change from baseline for each of these lesion counts both for those who completed up to the final visit 10 and to the end of study for the ITT population (i.e. all subjects dispensed treatment) using the models discussed above. Following apparent ICH guidelines, we emphasize the LOCF results in the ITT population (denoted "eos" in Table 4 below). Even though the centers were quite disparate in size, again the sponsor's results were quite similar.

Table 4. Study 404: End of Study Lesion Counts

Includes sample sizes, means and LS means, standard errors for each, and p-values from ANCOVA tests of treatment differences.

		Visit			
		10 (Completers)		eos (ITT)	
		Vehicle	Estrostep	Vehicle	Estrostrep
N		103	111	148	147
Inflammatory[1] Lesion Count	Mean	13.9	9.5	15.5	12.4
	Std Err	1.1	0.7	0.9	1.0
	LS Mean	14.3	9.8	15.5	12.2
	Std Err	0.8	0.8	0.9	0.9
	p-value*	0.0001		0.0082	
Total Comedones	Mean	29.8	21.0	29.9	23.9
	Std Err	1.9	1.4	1.5	1.5
	LS Mean	28.8	20.9	28.8	23.2
	Std Err	1.5	1.5	1.3	1.3
	p-value*	0.0001		0.0019	
Total Lesion Count[1]	Mean	43.8	30.5	45.3	36.3
	Std Err	2.7	1.8	2.0	2.1
	LS Mean	43.0	30.6	44.5	35.4
	Std Err	1.9	1.9	1.8	1.8
	p-value*	0.0001		0.0004	

Table 4. (cont.) Study 404: Lesion Counts

		Visit			
		10 (Completers)		eos (ITT)	
		Vehicle	Estrostep	Vehicle	Estrostrep
% Change in Inflam-[†] matory Lesions	Mean	52.2	67.4	45.5	58.1
	Std Err	3.5	2.5	3.0	3.0
	LS Mean	51.1	66.6	45.9	58.9
	Std Err	2.8	2.7	3.0	2.9
	p-value*	0.0001		0.0022	
% Change in Comedones	Mean	22.5	43.0	20.4	36.7
	Std Err	4.7	4.4	3.6	3.8
	LS Mean	24.4	44.2	23.9	38.2
	Std Err	4.4	4.3	3.5	3.5
	p-value*	0.0012		0.0038	
% Change in Total Lesions	Mean	37.6	56.1	33.0	48.1
	Std Err	3.2	2.5	2.7	2.5
	LS Mean	38.5	55.9	35.2	48.7
	Std Err	2.8	2.7	2.5	2.5
	p-value*	0.0001		0.0002	

*- p-value is for test of treatment differences in LS Means.

[†]- ANCOVA model uses baseline score with factors for site, treatment, and interaction. Otherwise interaction is deleted.

Note that all endpoints show statistically significant differences, both among the subjects who completed the study up to the final visit 10, and among the LOCF subjects at end of the study.

The following table, Text Table 5,shows the distribution of the investigator facial acne global evaluation in the Study 376-404. Again, recall that the protocol specifies that the primary endpoint for this measure is a dichotomization of this variable by those who were mild or better.

Table 5. Study 404: Investigator's Global Evaluation

	Visit							
	10(Completers)				eos (ITT)			
	Vehicle		Estrostep		Vehicle		Estrostep	
	n	%	n	%	n	%	n	%
Minimal	10	9.7	24	21.6	11	7.9	25	18.4
Mild	26	25.2	34	30.6	29	20.9	38	27.9
Mild to Moderate	36	35.0	34	30.6	50	36.0	43	31.6
Moderate	20	19.4	17	15.3	32	23.0	23	16.9
Marked	7	6.8	2	1.8	13	9.3	7	5.2
Severe	4	3.9	.	.	4	2.9	.	.

	Visit							
	10 (Completers)				eos (ITT)			
	Vehicle		Estrostep		Vehicle		Estrostep	
	n	%	n	%	n	%	n	%
Minimal or better	10	9.7	24	21.6	11	7.4	25	18.5
p-value*	0.009				0.012			
Mild or better	36	35.0	58	52.3	40	30.0	63	48.9
p-value*	0.004				0.001			
Mild/Moderate or worse	67	65.0	53	47.7	99	70.0	73	51.1

*p-value is for test of treatment differences in dichotomized response.

Thus, using the protocol specified primary binary endpoint, "mild or better", differences between Estrostep and its vehicle are statistically significant ($p \leq 0.0001$ at eos and $p \leq 0.0004$ at visit 10).

Finally, note from Appendix Tables A.14-A.16, when cast as closed tests, there is statistically significant evidence of efficacy by visit 5 or 6 (i.e. after 3-4 months of treatment).

Subgroup Analyses

Subject ages were used to define two subgroups, those subjects aged 19 or below, and those subjects age 20 and above. Appendix Table A.17 shows simple means and the associated standard errors for these age subgroups, as well as for race. In that table, one can note that for both age groups, all lesion counts show some superiority of Estrostep over vehicle. Similarly, but unlike the previous study, for all races, the lesion count scores show some superiority of Estrostep over vehicle.

Table A.18 in the Appendix gives the associated subgroup observed frequencies and proportions for the investigator's global evaluation. Again, note that for this global evaluation, for both age groups and all three race groups, Estrostep is generally superior to its vehicle.

VI. Adverse Events:

The sponsor provided tables of other adverse events during the study. Assessing adverse events is primarily a matter of clinical judgement, i.e., within the baliwick of the Medical Officer. Since the study was not designed for these tests, this should be interpreted as a "post hoc" analysis, with the possible problems associated with such post hoc analyses. However, it was felt that a multiplicity adjusted test of differences between Estrostep and its vehicle in the various adverse events might be useful. These analyses are based on the pooled adverse event data from both studies cited above.

To test the statistical significance of any differences in reported adverse events between Estrostep and its vehicle, the adverse events were first screened for those with five or more subjects experiencing the event. The number five was arbitrary, but reduces the number of adjustments required, and hence should increase power in the tests adjusted for multiplicity. Thirty-six adverse events met this criterion in the pooled data set. However, only the following were close to statistically significant (prior to adjusting for multiplicity of tests):

AE Code	Description	Incidence Vehicle	Estrostep	Unadjusted p-value	Adjusted p-value
0002B	Abdominal Pain	3/296	12/297	0.0329	0.5998
0354P	Nausea	9/296	19/297	0.0797	0.8890.
0847N	Bronchitis	12/296	4/297	0.0459	0.6928
1035K	Breast Pain	0/296	6/297	0.0305	0.5728
1076P	Metrorrhagia	8/296	53/297	0.0001	0.0001

The unadjusted p-value is the p-value from a Fisher Exact test of differences between Estrostep and its vehicle. All other unadjusted p-values were greater than .10. Adjusting the tests for the 36 potential comparisons using the techniques of Westfall and Young (1993) gives the "Adjusted p-value" cited above. In this particular case the adjustments were done using by sampling 5000 replicates from the permutation distribution of each table. These are used to approximate the distribution of the minimum p-value of all the tests. Unlike most other methods for correcting for multiplicity, features of the distribution and inter-test correlations are incorporated into the analysis.

Thus, we would conclude that even adjusting for the multiplicity of possible adverse events, metrorrhagia is statistically significantly much worse in the Estrostep group than in the corresponding vehicle group. Whether or not that observation is of practical significance again requires the clinical expertise of the Medical Officer.

REFERENCE:

Westfall, P.H. and Young, S.S. (1993) *Resampling-Based Multiple Testing: Examples and Methods for p-value Adjustment*, New York: John Wiley & Sons, Inc.

Conclusions (Which may be conveyed to the Sponsor):

1. In Study 376-403, treatment differences between the actual lesion counts and the percentage change from baseline at the end of the study were statistically significant for the ITT population (using LOCF imputation) at the end of the study both for comedones (original score $p \leq 0.0197$ and % change $p \leq 0.0187$) and total lesions (original score $p \leq 0.0118$ and % change $p \leq 0.0169$). The differences for inflammatory lesions in the ITT group were not quite statistically significant (though close) for both the original score and the percent change from baseline. These tests corresponded to an estimated 30% reduction from baseline in comedones in the Estrostep group versus a 17% reduction in the vehicle group (using least squares means), and an estimated 36% reduction from baseline in total lesions in the Estrostep group versus a 26% reduction in the vehicle group.

2. For the global facial evaluation in Study 376-403 there was an apparent trend favoring Estrostep over vehicle. However, the dichotomization of this score specified by the protocol was to define "success" as absent, minimal, or mild. While the trend favors Estrostep over vehicle (40% versus 30%), the difference was not strictly speaking, statistically significant at the usual level ($p \leq 0.083$). However, if the binary split had been done at minimal or better, the differences (16% versus 7%), the difference would have been statistically significant ($p \leq 0.015$), that is, results are sensitive to the specified cutpoint.

3. In the ITT population in Study 376-404 differences at the end of the study between vehicle and Estrostep were statistically significant for inflammatory lesions (original score $p \leq 0.0083$ and % change $p \leq 0.0022$), comedones (original score $p \leq 0.0014$ and % change $p \leq 0.0036$), and total lesions (original score $p \leq 0.0002$ and % change $p \leq 0.0001$). These tests corresponded to an estimated 59% reduction from baseline in inflammatory lesions in the Estrostep group versus a 46% reduction in the vehicle group. For comedones there was an estimated 38% reduction from baseline in the Estrostep group versus a 24% reduction in the vehicle group. For total lesions the estimated reduction was 49% in the Estrostep group versus a 35% reduction n the vehicle group.

4. For the global facial evaluation in Study 376-404 both the dichotomization of this score specified by the protocol (i.e., "success" is absent, minimal, or mild) or the more restrictive definition (i.e., "success" is absent or minimal) show statistically significant differences in favor of Estrostep ($p \leq 0.001$ or $p \leq 0.009$, respectively). These correspond to differences of 49% versus 30% or 8% versus 18%, respectively, in favor of Estrostep over vehicle.

5. Thus, the sponsor clearly "wins" in Study 376-404 study, with general, but not completely consistent support from Study 376-403. Study 376-403 did not show statistically significant differences on the dichotimized version of the physician's global facial acne evaluation as specified in the protocol. But the trends favored Estrostep, and there were statistically significant differences on a closely related endpoint in the ITT population at the end of the study. Note that these wins correspond to a difference of generally 10-20% or so in the reduction of lesions or improvement in global evaluation. Whether or not these moderately small effects are of clinical importance is a matter for the judgement of the Medical Officer, but from a statistical point of view there seems to be adequate evidence to support a claim of some efficacy.

6. Note that these studies were not designed to test differences in adverse events. But in an admittedly quite post hoc analysis, even adjusting for the multiplicity of possible adverse events, metrorrhagia was statistically significantly much worse in the Estrostep group (53/297) than in the

corresponding vehicle group (8/296). Whether or not that observation is of practical significance again requires the clinical expertise of the Medical Officer.

7. In the labelling (page 45) the sponsor claims that an improvement is apparent within 2-3 months after starting Estrostep. Even in the more favorable study, 376-404, Estrostep was not uniformly statistically significantly better than its vehicle until visit 7 (i.e. in the fifth month of the study), although it is close at visit 6 (month four). In Study 376-403, on the lesion count endpoints Estrostep was not statistically significantly better than placebo at any earlier time point. Thus this reviewer would not agree with the sponsor's claim.

Steve Thomson
Mathematical Statistician, Biometrics III

concur: **Mohamed Alosh, Ph.D.**
Team Leader, Biometrics III

cc:
Archival NDA: 21-276 Estrostep
HFD-540/Division File
HFD-540/Dr. Wilkin
HFD-540/Dr. Okun
HFD-540/Dr. Porres
HFD-540/Ms. Cintron
HFD-725/Dr. Huque
HFD-725/Dr. Alosh
HFD-725/Mr. Thomson
HFD-340/Dr. Lepay
This review has 19 pages including this signature page, plus a 21 page appendix.
Chron.

1 page(s) have been removed because it contains trade secret and/or confidential information that is not disclosable.

Appendix Table A.2. Study 376-403 and 376-404: Sensitivity Analysis of Pooling Example Significance Levels When Centers are Pooled Randomly

The following show example p-values from the ANCOVA tests of differences in treatment when those centers with less than 15 subjects are randomly pooled into two "new centers". The grouping labelled "cluster" corresponds to that used in the cluster analysis described in the report. The others correspond to random pooling of centers. The data are for the final, end-of-study ("eos") visit using last-observation-carried-forward technology. The models used (with or without interaction) are those used for the primary analysis as displayed in Text Tables 2 and 4.

Study 403

VISIT	Grouping	Inflammatory Lesions	Comedones	Total Lesions	% Chg in Inflam.	% Chg in Comedones	% Chg in Total
eos	cluster	0.05148	0.01967	0.01176	0.06448	0.01868	0.01691
eos	random01	0.05185	0.02219	0.01307	0.06335	0.02066	0.01843
eos	random02	0.05124	0.02242	0.01312	0.06220	0.02105	0.01843
eos	random03	0.05366	0.02145	0.01293	0.06547	0.02000	0.01839
eos	random04	0.05214	0.02179	0.01291	0.06342	0.01942	0.01771
eos	random05	0.05213	0.02426	0.01392	0.06325	0.02272	0.01955
eos	random06	0.05884	0.02478	0.01519	0.07140	0.02371	0.02105
eos	random07	0.05107	0.02254	0.01299	0.06204	0.02132	0.01835
eos	random08	0.05008	0.02275	0.01326	0.06191	0.02102	0.01851
eos	random09	0.05886	0.03031	0.01762	0.07143	0.02915	0.02494
eos	random10	0.05180	0.02378	0.01392	0.06362	0.02232	0.01973
eos	random11	0.05485	0.02258	0.01360	0.06721	0.02124	0.01913
eos	random12	0.06100	0.02886	0.01727	0.07405	0.02739	0.02433
eos	random13	0.05209	0.02109	0.01254	0.06511	0.02097	0.01847
eos	random14	0.05251	0.02142	0.01292	0.06354	0.01940	0.01738
eos	random15	0.06064	0.02791	0.01685	0.07381	0.02599	0.02348
eos	random16	0.06980	0.03037	0.01909	0.08513	0.02688	0.02565
eos	random17	0.05286	0.02383	0.01369	0.06417	0.02223	0.01908
eos	random18	0.05221	0.02252	0.01328	0.06380	0.02110	0.01864
eos	random19	0.05111	0.02117	0.01249	0.06280	0.01866	0.01682
eos	random20	0.04941	0.02224	0.01287	0.06090	0.02099	0.01831

Study 404

VISIT	SEQUENCE	Inflammatory Lesions	Comedones	Total Lesions	% Chg in Inflam.	% Chg in Comedones	% Chg in Total
eos	cluster	0.008149	.0018984	.0003579	.0022211	.0037563	.00015697
eos	random01	0.011928	.0022748	.0008303	.0028211	.0045077	.00018746
eos	random02	0.010739	.0019651	.0006697	.0034538	.0038763	.00016098
eos	random03	0.009821	.0020027	.0003992	.0032181	.0039247	.00016818
eos	random04	0.005641	.0019764	.0001635	.0019460	.0039344	.00016356
eos	random05	0.008656	.0020629	.0003998	.0030505	.0041366	.00017524
eos	random06	0.008509	.0019008	.0003961	.0023805	.0037661	.00015687
eos	random07	0.005801	.0019103	.0002245	.0018506	.0037729	.00015786
eos	random08	0.011161	.0021358	.0014357	.0024405	.0042515	.00017785
eos	random09	0.017660	.0018966	.0014598	.0058531	.0037697	.00015674
eos	random10	0.011233	.0019445	.0006204	.0032774	.0038519	.00015949
eos	random11	0.006679	.0018793	.0003068	.0019209	.0037158	.00015677
eos	random12	0.009964	.0020574	.0004894	.0028080	.0040732	.00016849
eos	random13	0.009703	.0019762	.0005320	.0032738	.0039231	.00016255
eos	random14	0.009221	.0019831	.0004556	.0027571	.0039726	.00016655
eos	random15	0.009347	.0019610	.0004695	.0024726	.0039557	.00016412
eos	random16	0.009775	.0019906	.0005895	.0030721	.0039267	.00016225
eos	random17	0.008446	.0019498	.0004101	.0023947	.0038471	.00016138
eos	random18	0.009262	.0020955	.0004274	.0027892	.0041485	.00017292
eos	random19	0.008518	.0019384	.0003542	.0023444	.0038606	.00016133
eos	random20	0.008531	.0018821	.0004073	.0022465	.0037421	.00015701

Note that these, and a couple of hundred other runs not displayed here, indicate that p-values seem to remain quite consistent across the various random assignments of sites to super-centers, and especially consistent with the center allocations defined by the cluster analysis above.

Appendix Table A.3. Study 376-403: Example Mixed Model Analyses

(Note the comments below applied to all lesion count analyses in both studies)

Inflammatory Lesions:

Covariance Parameter Estimates (REML)

Site	15.3
Site*Rx Grp	0.0
Visit*Site	1.9
Visit*Site*Rx Grp	2.3
Patient	44.4
Residual	53.0

The model was re-run with the three small variance components deleted. It made no practical difference. Specifying an unstructured covariance matrix also had no large effect.

Tests of Fixed Effects

Source	NDF	DDF	Type III F	Pr > F	
Inflam Baseline	1	262	91.41	0.0001	
Rx Grp	1	150	1.63	0.2037	
Visit	7	54	18.11	0.0001	Deleting the interaction term
Visit*Rx Grp	7	41.7	0.77	0.6133	made no difference.

Differences of Least Squares Means

Effect	Rx Grp	-	Rx Grp	Difference	Std Error	DF	t	Pr > |t|
Rx Grp	Estrostep		Vehicle	-1.26	0.98	150	-1.28	0.2037

Comedones:

Covariance Parameter Estimates (REML)

Site	112.6
Site*Rx Grp	0.0
Visit*Site	19.8
Visit*Site*Rx Grp	0.2
Patient	178.5
Residual	125.8

Tests of Fixed Effects

Source	NDF	DDF	Type III F	Pr > F
Comedone Baseline	1	278	119.67	0.0001
Rx Grp	1	192	4.23	0.0411
Visit	7	73.5	7.03	0.0001
Visit*Rx Grp	7	8.05	3.20	0.0619

Differences of Least Squares Means

Effect	Rx Grp	-	Rx Grp	Difference	Std Error	DF	t	Pr > |t|
Rx Grp	Estrostep		Vehicle	-3.64	1.77	192	-2.06	0.0411

Total Lesions:

Covariance Parameter Estimates (REML)

Site	180.8
Site*Rx Grp	0.0
Visit*Site	16.2
Visit*Site*Rx Grp	9.0
Patient	276.3
Residual	201.9

Tests of Fixed Effects

Source	NDF	DDF	Type III F	Pr > F
Total Lesion Baseline	1	277	110.47	0.0001
Rx Grp	1	156	4.41	0.0373
Visit	7	73.3	17.21	0.0001
Visit*Rx Grp	7	27.1	2.37	0.0501

Differences of Least Squares Means

Effect	Rx Grp	-	Rx Grp	Difference	Std Error	DF	t	Pr > |t|
Rx Grp	Estrostep		Vehicle	-4.88	2.32	156	-2.10	0.0373

Appendix Table A.4. Study 376-404: Example Mixed Model Analyses

Inflammatory Lesions: Covariance Parameter Estimates (REML)

Site	11.4
Site*Rx Grp	2.0
Visit*Site	0.6
Visit*Site*Rx Grp	0.0
Patient	42.9
Residual	44.8

Tests of Fixed Effects

Source	NDF	DDF	Type III F	Pr > F
Inflam. Baseline	1	254	86.04	0.0001
Rx Grp	1	15.5	7.61	0.0143
Visit	7	67.2	47.64	0.0001
Visit*Rx Grp	7	1459	5.90	0.0001

Differences of Least Squares Means

Effect	Rx Grp	- Rx Grp	Difference	Std Error	DF	t	Pr > |t|
Rx Grp	Estrostep	Vehicle	-2.89	1.05	15.3	-2.76	0.0143

Comedones: Covariance Parameter Estimates (REML)

Site	17.7
Site*Rx Grp	0.2
Visit*Site	4.5
Visit*Site*Rx Grp	0.3
Patient	101.0
Residual	104.0

Tests of Fixed Effects

Source	NDF	DDF	Type III F	Pr > F
Comedone Baseline	1	253	167.38	0.0001
Rx Grp	1	12.8	11.86	0.0044
Visit	7	86.3	14.53	0.0001
Visit*Rx Grp	7	62.3	4.04	0.0010

Differences of Least Squares Means

Effect	Rx Grp	- Rx Grp	Difference	Std Error	DF	t	Pr > |t|
Rx Grp	Estrostep	Vehicle	-4.70	1.37	12.8	-3.44	0.0044

Total Lesions: Covariance Parameter Estimates (REML)

Site	30.4
Site*Rx Grp	6.3
Visit*Site	5.7
Visit*Site*Rx Grp	0.0
Patient	178.8
Residual	163.6

Tests of Fixed Effects

Source	NDF	DDF	Type III F	Pr > F
Tot. Les. Baseline	1	247	163.61	0.0001
Rx Grp	1	13.9	13.10	0.0028
Visit	7	81	39.80	0.0001
Visit*Rx Grp	7	1465	7.79	0.0001

Differences of Least Squares Means

Effect	Rx Grp	- Rx Grp	Difference	Std Error	DF	t	Pr > |t|
Rx Grp	Estrostep	Vehicle	-7.49	2.07	13.9	-3.62	0.0028

Appendix Table A.5. Study 376-403: Association Between Lesion Counts and Global Evaluation

The following table displays the count of subjects who had at least a one unit improvement (i.e., a decrease) in the global evaluation from the first measurement (visit 3). Similarly the number of subjects with at least a two unit improvement, and the count of those with at least a three unit improvement in their global evaluation are also tabulated. These are cross-tabulated with the corresponding mean decrease in total lesion count for each of these decreasing subsets of subjects. N denotes the number of subjects with both a global evaluation and lesion counts at that visit. Results for the pooled super centers are listed under Pooled 1 and Pooled 2 below. Finally a summary over all centers is denoted by the listing "All." Note this table was provided at the request of the Medical Officer and is meant to display the association between the global change and lesion counts. A more analytic approach to this association is given in section II.D (pages 8-9).

Study 403

VISIT		04		05		06		07		08		09		10		eos	
Center		Veh.	Es.	Veh.	Es.	Veh.	Es.	Veh.	Es.	Veh.	Es.	Veh.	Es.	Veh.	Es.	Veh.	Es.
006	N	6	5	6	4	4	5	5	5	5	5	5	5	5	5	7	6
Glb chg>=1		3	3	2	2	3	3	4	3	3	4	2	3	4	4	4	4
Tot. Les.		8.7	15.7	26.5	30.5	-9.7	28.3	32.3	29.3	29.3	44.8	31.0	53.7	17.8	54.5	17.8	54.5
Glb chg>=2		0	1	0	2	0	2	2	1	2	2	1	2	1	3	1	3
Tot. Les.		.	16.0	.	30.5	.	25.5	52.0	33.0	40.0	34.5	31.0	68.5	25.0	62.3	25.0	62.3
Glb chg>=3		0	0	0	0	0	0	1	1	0	1	0	1	0	1	0	1
Tot. Ls.		.	.	.	.	.	.	51.0	33.0	.	33.0	.	33.0	.	33.0	.	33.0
009	N	61	64	56	61	52	62	46	55	42	52	40	51	41	51	62	66
Glb chg>=1		26	28	23	28	29	41	28	35	23	36	23	35	24	34	33	41
Tot. Les.		11.2	7.5	16.7	16.8	22.5	17.7	23.4	25.1	26.3	27.0	20.5	26.3	24.2	27.2	24.2	26.8
Glb chg>=2		5	6	9	8	4	12	9	15	9	16	9	16	9	14	12	15
Tot. Les.		15.4	12.2	22.0	26.6	18.5	31.0	32.6	35.0	35.1	32.5	28.0	31.4	34.9	33.4	33.8	31.2
Glb chg>=3		0	0	1	1	0	2	3	4	1	5	2	4	2	4	3	5
Tot. Les.		.	.	28.0	37.0	.	31.5	44.0	44.0	8.0	48.2	17.0	36.8	26.0	44.0	37.0	35.4
010	N	12	12	12	11	12	11	12	10	10	9	12	10	12	10	15	15
Glb chg>=1		2	2	4	2	4	7	5	8	4	7	8	6	7	8	7	9
Tot. Les.		5.0	33.0	9.5	4.0	-7.3	3.9	-1.4	3.4	2.5	19.1	-11.3	21.2	-3.4	12.8	-3.4	14.8
Glb chg>=2		0	1	1	0	1	0	0	1	3	2	1	2	3	3	3	3
Tot. Les.		.	29.0	-4.0	.	-30.0	.	.	-37.0	-7.0	-8.5	0.0	22.0	23.3	15.3	23.3	15.3
012	N	7	6	7	6	6	5	6	4	6	4	6	3	6	4	8	8
Glb chg>=1		0	2	1	2	0	2	1	1	1	2	0	1	0	2	1	2
Tot. Les.		.	28.5	3.0	33.0	.	53.0	7.0	22.0	11.0	91.0	.	112	.	81.0	5.0	81.0
Glb chg>=2		0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
Tot. Les.		.	.	.	.	.	.	.	.	.	.	.	.	.	100	.	100

Appendix Table A.5 (cont.): Study 376-403

Study 403

VISIT		04		05		06		07		08		09		10		eos	
Center		Veh.	Es.	Veh.	Es.	Veh.	Es.	Veh.	Es.	Veh.	Es.	Veh.	Es.	Veh.	Es.	Veh.	Es.
014	N	10	13	9	12	8	10	7	9	5	8	5	6	5	6	14	13
Glb chg>=1		4	5	5	7	4	7	3	5	4	7	3	5	3	5	6	10
Tot. Les.		36.8	29.8	40.4	33.1	32.0	44.4	40.0	52.4	41.0	50.9	42.7	54.8	44.0	45.8	46.2	38.3
Glb chg>=2		2	0	3	1	2	3	1	4	2	3	2	4	3	3	4	4
Tot. Les.		51.5	.	54.0	48.0	36.0	39.3	76.0	54.8	55.5	59.3	61.0	60.0	44.0	63.3	41.8	51.3
Glb chg>=3		0	0	1	0	1	0	1	0	2	0	2	0	1	0	1	0
Tot. Les.		.	.	61.0	.	54.0	.	76.0	.	55.5	.	61.0	.	76.0	.	76.0	.
Pooled 1	N	11	14	10	13	7	10	6	10	6	11	5	10	6	11	13	16
Glb chg>=1		2	5	6	5	4	5	3	4	3	7	3	6	4	8	8	9
Tot. Les.		2.5	25.0	34.2	31.0	42.3	28.8	43.0	27.8	28.0	30.1	60.7	41.7	30.8	33.0	28.8	28.8
Glb chg>=2		0	1	1	1	1	0	0	1	0	2	0	3	0	3	2	3
Tot. Les.		.	24.0	40.0	48.0	70.0	.	.	78.0	.	61.0	.	48.3	.	56.3	54.5	56.3
Pooled 2	N	18	15	16	15	17	14	11	13	11	13	10	13	11	13	20	17
Glb chg>=1		4	5	4	6	5	6	3	8	3	9	5	10	5	9	8	10
Tot. Les.		7.3	20.2	17.8	10.7	15.0	16.7	25.7	9.4	27.0	26.2	6.8	26.5	12.0	31.2	15.3	27.2
Glb chg>=2		3	0	2	1	2	3	1	3	3	5	0	4	0	7	2	8
Tot. Les.		5.3	.	25.0	8.0	32.5	11.0	31.0	-1.3	27.0	27.8	.	28.0	.	28.0	32.0	23.4
Glb chg>=3		0	0	1	0	1	0	0	0	0	0	0	1	0	2	1	2
Tot. Les.		.	.	53.0	.	31.0	.	.	.	.	.	.	9.0	.	19.0	31.0	19.0
ALL	N	125	129	116	122	106	117	93	106	85	102	83	98	86	100	139	141
Glb chg>=1		41	50	45	52	49	71	47	64	41	72	44	66	47	70	67	85
Tot. Les.		12.4	15.1	21.2	20.3	19.7	21.1	23.6	22.8	25.4	31.5	7.9	32.0	20.0	31.2	22.1	29.7
Glb chg>=2		10	9	16	13	10	20	13	25	19	30	13	31	16	34	24	37
Tot. Les.		19.6	15.8	27.9	29.1	25.1	28.7	38.8	32.6	29.8	33.7	31.2	38.1	33.8	39.9	35.0	36.8
Glb chg>=3		0	0	3	1	2	2	5	5	3	6	4	6	3	7	5	8
Tot. Les.		.	.	47.3	37.0	42.5	31.5	51.8	41.8	39.7	45.7	39.0	31.5	42.7	35.3	43.6	31.0

APPEARS THIS WAY
ON ORIGINAL

Appendix Table A.6. Study 376-404: Association Between Lesion Counts and Global Evaluation

The table is constructed similar to the preceding table.

Study 404

VISIT		04		05		06		07		08		09		10		eos	
Center		Veh.	Es.	Veh.	Es.	Veh.	Es.	Veh.	Es.	Veh.	Es.	Veh.	Es.	Veh.	Es.	Veh.	Es.
001	N	9	10	9	10	9	9	9	9	9	9	9	9	9	9	9	11
	Glb chg>=1	3	2	3	5	6	4	4	6	6	5	6	6	6	7	6	9
	Tot. Les.	18.0	4.5	4.7	9.6	16.0	19.5	14.0	18.2	10.2	20.4	23.5	22.0	28.5	21.0	28.5	28.0
	Glb chg>=2	0	0	0	0	1	0	0	0	1	1	0	0	1	2	1	3
	Tot. Les.	.	.	.	.	25.0	.	.	.	14.0	27.0	.	.	39.0	21.0	39.0	33.0
	Glb chg>=3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	Tot. Les.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	57.0
003	N	18	15	16	14	16	13	13	12	13	13	14	12	14	13	18	17
	Glb chg>=1	4	1	5	5	6	6	5	6	4	8	6	8	8	7	9	7
	Tot. Les.	3.3	11.0	11.0	24.4	13.5	29.8	28.4	21.0	28.5	20.5	24.2	23.5	26.5	19.1	25.8	19.1
	Glb chg>=2	1	0	3	1	2	2	2	2	2	0	4	1	4	1	5	1
	Tot. Les.	23.0	.	14.7	23.0	21.0	38.5	39.5	31.5	23.0	.	19.3	25.0	29.5	25.0	27.6	25.0
	Glb chg>=3	0	0	0	0	0	0	2	0	0	0	0	0	1	0	1	0
	Tot. Les.	.	.	.	.	.	.	39.5	.	.	.	.	.	65.0	.	65.0	.
005	N	7	9	6	8	6	8	6	8	5	8	4	8	5	8	7	10
	Glb chg>=1	1	1	0	2	1	1	1	2	2	3	2	2	2	4	2	4
	Tot. Les.	7.0	20.0	.	5.0	5.0	18.0	8.0	13.0	5.5	16.0	8.5	15.0	8.0	26.0	8.0	26.0
	Glb chg>=2	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
	Tot. Les.	.	.	.	.	.	.	.	.	.	.	7.0	.	.	.	.	.
007	N	13	13	12	12	12	13	12	13	11	13	11	13	11	13	14	15
	Glb chg>=1	1	2	3	3	2	6	3	5	6	8	3	8	5	10	7	10
	Tot. Les.	28.0	27.5	7.7	21.0	42.5	43.8	45.3	37.8	36.5	50.5	38.3	43.1	46.6	49.7	44.6	49.7
	Glb chg>=2	0	1	0	1	0	2	1	0	2	2	1	3	1	5	1	5
	Tot. Les.	.	43.0	.	50.0	.	64.5	24.0	.	53.5	54.0	28.0	36.7	55.0	42.2	28.0	42.2
	Glb chg>=3	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
	Tot. Les.	.	.	.	.	.	.	.	.	34.0	58.0	28.0	28.0	55.0	56.0	28.0	56.0
010	N	19	21	20	21	16	21	12	18	14	15	13	18	15	19	23	23
	Glb chg>=1	8	7	8	7	6	12	5	13	5	13	7	14	6	12	7	13
	Tot. Les.	16.0	20.1	25.4	25.6	17.2	23.3	28.6	25.8	26.6	21.2	19.4	24.6	28.0	27.2	23.7	26.1
	Glb chg>=2	1	1	3	1	1	2	2	5	3	3	2	3	3	5	3	5
	Tot. Les.	37.0	32.0	33.7	49.0	51.0	28.0	42.0	30.0	28.7	27.0	46.5	16.7	39.0	32.2	39.0	32.2
	Glb chg>=3	1	0	1	0	1	0	1	0	0	0	1	0	2	0	2	0
	Tot. Les.	37.0	.	45.0	.	51.0	.	49.0	.	.	.	53.0	.	45.0	.	45.0	.
011	N	10	11	10	11	10	11	10	11	9	10	9	11	9	11	11	12
	Glb chg>=1	2	1	3	2	3	4	3	6	3	9	4	11	4	11	5	11
	Tot. Les.	11.0	9.0	17.0	15.5	24.3	22.3	26.3	26.0	30.0	28.4	23.3	35.9	30.5	37.4	27.6	37.4
	Glb chg>=2	0	0	0	0	2	1	0	2	0	4	0	6	1	8	1	8
	Tot. Les.	.	.	.	.	23.5	19.0	.	35.5	.	36.8	.	40.2	30.0	43.3	30.0	43.3
	Glb chg>=3	0	0	0	0	0	0	0	1	0	2	0	2	0	3	0	3
	Tot. Les.	.	.	.	.	.	.	.	48.0	.	41.5	.	45.5	.	39.0	.	39.0

Appendix Table A.6. (cont.): Study 376-404

Study 404

VISIT		04		05		06		07		08		09		10		eos	
012	N	15	12	13	11	12	11	11	11	10	10	10	10	10	15	14	
Glb chg>=1		5	3	6	5	7	5	7	6	7	8	8	7	7	10	8	11
Tot. Les.		10.0	10.0	11.0	5.4	15.0	2.6	18.7	9.3	10.1	14.0	14.9	12.0	17.1	20.2	18.3	20.2
Glb chg>=2		0	0	0	1	1	2	0	1	1	1	1	1	1	1	1	1
Tot. Les.		.	.	.	11.0	28.0	3.5	.	-11.0	34.0	5.0	34.0	-11.0	19.0	33.0	19.0	33.0
014	N	12	11	12	11	11	10	11	10	10	10	9	10	11	10	12	13
Glb chg>=1		4	5	4	4	5	5	6	7	5	6	7	7	6	6	6	7
Tot. Les.		6.8	12.8	17.8	22.3	27.8	32.6	20.3	34.9	26.2	30.0	26.1	38.0	29.3	36.8	29.3	33.4
Glb chg>=2		0	0	0	1	2	1	3	3	3	5	2	4	5	4	5	4
Tot. Les.		.	.	.	71.0	34.0	80.0	26.3	48.7	16.0	32.8	30.5	33.8	35.4	31.5	35.4	31.5
Glb chg>=3		0	0	0	0	1	1	1	1	1	1	1	1	4	1	4	1
Tot. Les.		.	.	.	.	45.0	80.0	49.0	83.0	19.0	70.0	28.0	55.0	40.5	71.0	40.5	71.0
Pooled 1 N		14	9	12	8	11	8	11	8	10	6	8	7	11	8	15	12
Glb chg>=1		5	3	3	3	6	6	7	6	5	6	4	5	6	7	8	9
Tot. Les.		6.0	18.7	26.3	25.0	3.0	24.5	14.4	25.2	17.2	27.7	21.8	31.4	10.2	31.3	9.3	24.7
Glb chg>=2		2	2	1	2	2	3	2	3	1	2	1	3	1	4	2	5
Tot. Les.		21.5	14.5	54.0	30.0	2.0	27.0	20.0	40.7	46.0	53.5	58.0	39.0	62.0	33.8	35.5	29.2
Glb chg>=3		0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	2
Tot. Les.		.	.	.	.	.	.	43.0	45.0	46.0	46.0	58.0	47.0	62.0	46.0	62.0	28.5
Pooled 2 N		10	11	10	10	9	10	7	10	9	6	7	7	8	10	16	14
Glb chg>=1		3	4	3	5	3	6	4	6	2	3	5	3	5	7	7	8
Tot. Les.		16.3	10.5	1.3	28.0	11.7	25.2	29.0	25.5	39.5	37.0	30.2	26.7	29.6	31.1	25.3	31.9
Glb chg>=2		0	1	0	2	0	3	1	3	0	2	1	1	0	3	0	3
Tot. Les.		.	4.0	.	26.0	.	36.3	70.0	27.7	.	39.5	59.0	45.0	.	36.3	.	36.3
Glb chg>=3		0	0	0	1	0	1	0	1	0	0	0	1	0	1	0	1
Tot. Les.		.	.	.	36.0	.	36.0	.	32.0	.	.	.	45.0	.	40.0	.	40.0
ALL	N	127	122	120	116	112	114	102	110	100	100	94	105	103	111	140	141
Glb chg>=1		36	29	38	41	45	55	45	63	45	69	52	71	55	81	65	89
Tot. Les.		11.3	15.1	14.9	19.1	16.4	25.1	23.0	24.5	22.1	26.3	22.8	28.5	25.9	30.6	24.7	30.0
Glb chg>=2		4	5	7	9	11	16	11	19	13	20	13	22	17	33	19	35
Tot. Les.		25.8	21.6	28.4	35.1	24.1	34.9	34.2	32.8	29.3	35.9	32.1	32.4	34.7	36.0	32.6	35.9
Glb chg>=3		1	0	1	1	2	2	5	4	3	5	4	6	9	7	9	9
Tot. Les.		37.0	.	45.0	36.0	48.0	58.0	44.0	52.0	33.0	51.4	41.8	48.8	45.2	47.1	45.2	44.2

APPEARS THIS WAY
ON ORIGINAL

Appendix Table A.7. Study 376-403: Patient Demographics

	Vehicle	Estrostep
No. Enrolled	148	150
No. Completed	86	102
No. Discontinued:	62	48
Adverse Event	4	13
Lack of Efficacy	8	3
Lack of Compliance	4	3
Pregnancy	5	0
Other	41	29
Mean Age (Range)	24.0 (14-45)	25.0 (13-48)
Race: White/Caucasian	98 (66%)	100 (67%)
Black	25 (17%)	18 (12%)
Asian	5 (3%)	5 (3%)
Hispanic	17 (11%)	20 (13%)
Other	3 (2%)	7 (5%)

Appendix Table A.8. Study 376-404: Patient Demographics

	Vehicle	Estrostep
No. Enrolled	148	147
No. Completed	104	111
No. Discontinued:	44	36
Adverse Event	3	7
Lack of Efficacy	1	5
Lack of Compliance	8	4
Pregnancy	5	1
Other	27	19
Mean Age (Range)	23.9 (14-48)	23.6 (14-48)
Race: White/Caucasian	104 (71%)	102 (70%)
Black	21 (14%)	20 (14%)
Asian	2 (1%)	13 (9%)
Hispanic	17 (12%)	10 (7%)
Other	3 (2%)	1 (1%)

Although not displayed in any output included with this report, the tests of the effect of treatment on the probability of being a dropout were statistically very non-significant in both studies.

Appendix Table A.9. Study 376-403: Response Profiles in Lesion Counts: Completers

This table displays the least squares means of treatment effects from an ANOVA model with site, treatment, and interaction as factors and baseline as a covariate for those subjects with data at that time (i.e. no imputation, except at the eos measure). This is designed to indicate the overall means over time. In addition a each time point there is a p-value corresponding to the test of no mean differences between treatment at that visit. As noted in the section II.C, to control the Type I error associated with analyzing multiple visits for each lesion count or global assessment, these should be read as a closed family of tests. That is, the results at week 9 are evaluated if and only if the test of difference between vehicle and Estrostep are statistically significant at week 10. Similarly the week 8 results are evaluated if and only if the test of difference between vehicle and Estrostep are statistically significant at week 9. If at any visit the differences are not statistically significant, the analysis stops and reports any observed statistical significance up to that time point.

	VISIT																	
	03		04		05		06		07		08		09		10		eos	
	Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est
N	135	134	125	130	117	123	106	117	93	106	85	102	83	98	86	100	148	149
Inflammatory																		
LS Mean	23.0	24.6	21.9	20.5	18.3	19.4	19.7	15.7	15.5	15.0	13.6	13.0	14.0	13.8	14.8	13.4	18.9	16.2
Std Err	1.1	1.1	1.1	1.2	1.2	1.3	1.2	1.1	1.3	1.3	1.4	1.4	1.4	1.4	1.4	1.4	1.1	1.1
p-value		.2993		.3901		.5436		.0162		.7947		.7370		.9021		.4802		.0934
Comedones																		
LS Mean	39.0	40.0	38.4	33.0	34.8	31.6	34.7	31.9	32.6	33.2	31.1	26.3	32.7	23.2	30.0	25.6	36.4	30.7
Std Err	1.8	1.9	2.1	2.2	2.1	2.2	2.4	2.4	2.6	2.6	2.3	2.2	2.5	2.6	2.4	2.4	2.0	2.1
p-value		.7090		.0769		.2885		.3961		.8558		.1351		.0087		.1935		.0470
Total Lesions																		
LS Mean	61.9	64.4	60.3	53.5	53.1	51.1	54.5	47.8	48.1	48.1	44.8	39.1	46.6	36.9	44.8	38.8	55.3	46.8
Std Err	2.2	2.3	2.6	2.6	2.5	2.6	2.9	2.8	3.3	3.3	3.0	3.0	3.1	3.3	3.2	3.2	2.7	2.8
p-value		.4402		.0673		.5837		.1028		.9919		.1797		.0330		.1864		.0303
% Change in Inflammatory Lesions																		
LS Mean	21.1	15.6	25.2	27.9	37.8	31.1	34.7	43.8	47.2	48.7	52.7	53.9	51.4	51.8	47.5	53.2	32.5	42.8
Std Err	3.7	3.8	4.0	4.1	4.3	4.4	4.2	4.0	4.6	4.6	5.0	4.8	5.0	5.1	5.0	5.0	4.0	4.1
p-value		.3048		.6451		.2806		.1201		.8159		.8558		.9509		.4179		.0754
% Change in Comedones																		
LS Mean	12.8	11.1	14.3	20.6	22.2	23.8	21.3	23.4	25.7	19.6	32.0	35.2	25.5	41.1	31.9	38.9	16.9	30.4
Std Err	4.5	4.7	5.1	5.3	4.9	5.1	5.5	5.4	6.3	6.4	5.5	5.4	5.9	6.1	6.1	6.1	5.1	5.2
p-value		.7916		.3911		.8255		.7864		.4956		.6840		.0676		.4159		.0617
% Change in Total Lesions																		
LS Mean	16.8	14.2	19.9	24.3	29.6	28.7	27.2	32.2	35.7	33.1	41.3	45.3	37.0	48.2	39.9	46.4	25.5	36.5
Std Err	3.2	3.3	3.6	3.7	3.4	3.5	4.0	3.9	4.5	4.5	4.1	4.0	4.2	4.4	4.3	4.4	3.7	3.8
p-value		.5773		.3883		.8461		.3686		.6895		.4842		.0652		.2938		.0396

Appendix Table A.10. Study 376-403: Response Profiles in Global Acne Assessment : Completers

To maintain overall error p-values are to be read from the right to the left, stopping at the first nonsignificant p-value.

	Visit 03		04		05		06		07		08		09		10	
	0Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est
1	.	.	.	.	.	.	.	.	.	1	.	2	.	1	.	2
%	.	.	.	.	.	.	.	.	.	0.9	.	2.0	.	1.0	.	2.0
2	1	2	2	4	4	5	3	8	10	9	3	13	4	19	7	19
%	0.7	1.5	1.6	3.1	3.4	4.1	2.8	6.8	10.8	8.5	3.5	12.7	4.8	19.4	8.1	19.0
3	13	14	22	22	26	25	24	29	21	27	30	30	25	21	25	24
%	9.6	10.4	17.6	17.1	22.4	20.3	22.6	24.8	22.6	25.5	35.3	29.4	30.1	21.4	29.1	24.0
4	47	43	46	49	42	52	40	55	29	47	29	42	27	39	33	38
%	34.8	32.1	36.8	38.0	36.2	42.3	37.7	47.0	31.2	44.3	34.1	41.2	32.5	39.8	38.4	38.0
5	63	68	46	52	38	33	37	21	32	21	20	14	25	17	19	14
%	46.7	50.7	36.8	40.3	32.8	26.8	34.9	17.9	34.4	19.8	23.5	13.7	30.1	17.3	22.1	14.0
6	11	7	9	2	6	8	2	4	1	1	3	1	2	1	2	3
%	8.1	5.2	7.2	1.6	5.2	6.5	1.9	3.4	1.1	0.9	3.5	1.0	2.4	1.0	2.3	3.0
N	135	134	125	129	116	122	106	117	93	106	85	102	83	98	86	100
Global=1,2																
n	1	2	2	4	4	5	3	8	10	10	3	15	4	20	7	21
%	0.7	1.5	1.6	3.1	3.4	4.1	2.8	6.8	10.8	9.4	3.5	14.7	4.8	20.4	8.1	21.0
p-value	.467		.465		.771		.198		.644		.013		.004		.019	
Global=1,2,3																
n	14	16	24	26	30	30	27	37	31	37	33	45	29	41	32	45
%	10.4	11.9	19.2	20.2	25.9	24.4	25.5	31.6	33.3	34.9	38.8	44.1	34.9	41.8	37.2	45.0
p-value	.647		.857		.803		.394		.929		.579		.476		.345	
-Global>3																
GLOB47	121	118	101	103	86	92	79	80	62	69	52	57	54	57	54	55

APPEARS THIS WAY
ON ORIGINAL

Appendix Table A.11. Study 376-403: Response Profiles: LOCF subjects

To maintain overall error p-values are to be read from the right to the left, stopping at the first nonsignificant p-value.

	Visit 03		04		05		06		07		08		09		10	
	Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est
N	148	149	148	149	148	149	148	149	148	149	148	149	148	149	149	149
Inflammatory																
LS Mean	23.8	24.8	22.7	21.3	20.2	19.9	21.3	17.9	19.0	17.0	18.0	16.2	18.2	16.2	18.9	16.2
Std Err	1.0	1.0	1.0	1.0	1.1	1.1	1.0	1.0	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1
p-value	.4829		.3264		.8515		.0212		.1960		.2427		.2111		.0934	
Comedones																
LS Mean	39.8	41.2	40.3	36.4	38.7	35.1	39.1	34.7	38.1	35.0	37.2	31.4	38.3	29.4	36.4	30.7
Std Err	1.7	1.8	1.9	2.0	2.0	2.0	2.0	2.1	2.1	2.2	2.0	2.0	2.0	2.1	2.0	2.1
p-value	.5777		.1591		.2074		.1303		.3026		.0380		.0022		.0470	
Total Lesions																
LS Mean	63.6	65.8	63.0	57.6	58.9	55.1	60.4	52.7	57.1	51.9	55.2	47.5	56.5	45.7	55.3	46.8
Std Err	2.1	2.1	2.3	2.4	2.5	2.5	2.5	2.6	2.8	2.8	2.6	2.7	2.7	2.7	2.7	2.8
p-value	.4586		.1128		.2813		.0359		.1860		.0380		.0050		.0303	
% Change in Inflammatory																
LS Mean	18.1	14.5	22.4	25.6	30.0	29.1	26.8	35.8	32.8	40.1	36.2	42.7	35.9	42.2	32.5	42.8
Std Err	3.4	3.5	3.6	3.6	3.9	3.9	3.7	3.7	3.9	4.0	4.0	4.1	4.0	4.1	4.0	4.1
p-value	.4541		.5282		.8666		.0871		.1919		.2611		.2727		.0754	
% Change in Comedones																
LS Mean	10.7	9.2	9.2	17.0	13.0	19.8	11.1	21.0	13.1	20.1	17.2	28.1	13.1	32.1	16.9	30.4
Std Err	4.2	4.4	4.6	4.7	4.8	4.9	4.8	4.9	5.2	5.4	4.8	4.9	4.9	5.0	5.1	5.2
p-value	.7975		.2385		.3169		.1468		.3505		.1169		.0066		.0617	
% Change in Total Lesions																
LS Mean	14.4	12.6	15.7	21.0	21.2	24.9	18.8	27.7	22.9	29.6	26.6	35.6	23.7	37.7	25.5	36.5
Std Err	3.0	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	3.6	3.7	3.6	3.7	3.7	3.8
p-value	.6774		.2532		.4454		.0814		.2225		.0839		.0076		.0396	

	Visit 03		04		05		06		07		08		09		10	
	Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est
N	135	134	135	135	135	135	135	135	135	135	135	135	135	135	135	135
Global=1,2																
n	1	2	2	5	4	6	4	9	12	12	6	17	8	23	10	25
%	0.7	1.5	1.5	3.7	3.0	4.4	3.0	6.7	8.9	8.9	4.4	12.6	5.9	17.0	7.4	18.5
p-value	.467		.243		.543		.152		.985		.015		.004		.005	
Global=1,2,3																
n	14	16	24	28	32	36	30	44	37	49	41	58	41	56	42	60
%	10.4	11.9	17.8	20.7	23.7	26.7	22.2	32.6	27.4	36.3	30.4	43.0	30.4	41.5	31.1	44.4
p-value	.647		.521		.553		.053		.107		.026		.050		.020	
Global>3																
n	121	118	111	107	103	99	105	91	98	86	94	77	94	79	93	75

Appendix Table A.12. Study 403: Subgroup Lesion Counts

Visit		10 (Completers)				eos (ITT)			
Age Group		19-		20+		19-		20+	
		Veh	Est	Veh	Est	Veh	Est	Veh	Est
Inflammatory Lesions	N	27	36	59	64	48	45	100	104
	Mean	21.5	16.3	14.6	14.9	23.1	17.2	17.2	16.2
	Std Error	2.4	1.7	1.3	1.4	1.9	1.7	1.2	1.1
Comedones	Mean	33.4	29.6	28.8	25.5	40.7	33.8	31.3	28.4
	Std Error	3.9	3.5	3.0	3.3	4.0	3.6	2.3	2.4
Total Lesions	Mean	54.9	45.9	43.4	40.4	63.8	51.1	48.4	44.6
	Std Error	5.7	4.5	3.8	4.1	5.3	4.6	3.1	2.9
% Change in Inflam. Lesions	Mean	28.7	49.8	47.0	44.9	24.6	44.7	37.5	39.8
	Std Error	8.4	4.9	5.0	5.0	6.5	5.5	4.0	4.0
% Change in Comedones	Mean	26.5	44.5	28.1	39.8	19.0	37.5	20.6	33.3
	Std Error	8.0	5.7	7.1	7.7	6.3	6.0	6.0	5.3
% Change in Total Lesions	Mean	29.8	47.4	38.1	43.1	24.6	41.6	29.1	36.7
	Std Error	6.3	4.5	4.7	5.4	4.6	4.7	4.3	3.9

Study 403

Visit		10 (Completers)						eos (ITT)					
Race		White		Black		Other		White		Black		Other	
		Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est
	N	58	75	15	10	13	15	98	100	25	18	25	31
Inflammatory Lesions	Mean	19.9	15.8	10.2	14.0	10.5	14.3	22.1	16.4	12.7	16.4	13.8	17.1
	Std Error	1.5	1.2	2.2	4.1	1.9	3.3	1.4	1.1	1.6	2.7	1.5	2.3
Comedones	Mean	30.1	27.3	34.9	29.4	25.6	23.6	34.7	29.3	37.6	38.9	29.6	27.3
	Std Error	2.9	2.7	7.4	10.9	3.4	5.7	2.6	2.3	6.0	8.2	3.5	3.7
Total Lesions	Mean	50.0	43.1	45.1	43.4	36.1	37.9	56.8	45.7	50.2	55.3	43.4	44.3
	Std Error	4.0	3.4	9.0	14.5	4.2	7.0	3.6	2.8	7.0	10.5	4.4	4.8
% Change in Inflam. Lesions	Mean	29.5	46.1	65.9	41.1	65.0	52.8	23.9	44.0	53.9	32.4	49.8	37.9
	Std Error	5.6	3.9	6.5	16.6	5.0	9.9	4.5	3.4	6.0	11.8	6.1	8.9
% Change in Comedones	Mean	23.9	40.8	33.6	44.6	37.4	42.5	16.8	33.7	23.7	38.3	29.3	35.1
	Std Error	7.4	6.4	11.0	15.4	6.9	12.6	6.1	5.5	10.5	9.5	6.4	7.5
% Change in Total Lesions	Mean	29.9	43.9	45.6	42.8	49.0	49.8	23.1	38.6	35.2	36.4	37.9	37.7
	Std Error	4.9	4.5	8.3	15.3	5.1	7.7	4.3	3.8	7.5	9.6	5.7	6.2

APPEARS THIS WAY
ON ORIGINAL

Appendix Table A.13. Study 403: Subgroup Global Evaluations

Visit			10 (Completers)				eos (ITT)			
Age Group			19-		20+		19-		20+	
			Veh	Est	Veh	Est	Veh	Est	Veh	Est
1. Absent	n		.	1	.	1	.	1	.	1
	%		.	2.8	.	1.5	.	2.3	.	1.0
2. Minimal	n		.	4	7	15	2	4	8	17
	%		.	11.1	11.9	23.1	4.4	9.1	8.5	17.3
3. Mild	n		7	5	18	20	7	5	25	29
	%		25.9	13.9	30.5	30.8	15.6	11.4	26.6	29.6
4. Mild to Moderate	n		10	19	23	19	16	23	34	29
	%		37.0	52.8	39.0	29.2	35.6	52.3	36.2	29.6
5. Moderate	n		8	7	11	7	18	10	22	16
	%		29.6	19.4	18.6	10.8	40.0	22.7	23.4	16.3
6. Marked	n		2	.	.	3	2	1	5	5
	%		7.4	.	.	4.6	4.4	2.3	5.3	5.1
7. Severe7	n		.	.	.	.	.	.	.	1
	%		.	.	.	.	.	.	.	1.0
Grouped 1		%	0.0	2.8	0.0	1.5	0.0	2.3	0.0	1.0
1,2		%	0.0	13.9	11.9	24.6	4.4	11.4	8.5	18.4
1-3		%	25.9	27.8	42.4	55.4	20.0	22.7	35.1	48.0

Study 403

Visit			10 (Completers)						eos (ITT)					
Race			White		Black		Other		White		Black		Other	
			Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est
1. Absent	n		.	2	.	.	.	.	.	2	.	.	.	.
	%		.	2.7	.	.	.	.	.	2.1	.	.	.	.
2. Minimal	n		4	12	2	3	1	4	5	14	2	3	3	4
	%		6.9	16.0	13.3	30.0	7.7	25.0	5.4	14.4	8.7	17.6	13.0	14.3
3. Mild	n		14	20	7	2	4	3	17	24	9	4	6	6
	%		24.1	26.7	46.7	20.0	30.8	18.8	18.3	24.7	39.1	23.5	26.1	21.4
4. Mild to Moderate	n		23	30	3	3	7	5	35	38	6	6	9	8
	%		39.7	40.0	20.0	30.0	53.8	31.3	37.6	39.2	26.1	35.3	39.1	28.6
5 Moderate	n		15	11	3	1	1	2	30	17	6	3	4	6
	%		25.9	14.7	20.0	10.0	7.7	12.5	32.3	17.5	26.1	17.6	17.4	21.4
6 Marked	n		2	.	.	1	.	2	6	2	.	1	1	3
	%		3.4	.	.	10.0	.	12.5	6.5	2.1	.	5.9	4.3	10.7
7 Severe	n		.	.	.	.	.	.	.	.	.	.	.	1
	%		.	.	.	.	.	.	.	.	.	.	.	3.6
Grouped 1	%		0.0	2.7	0.0	0.0	0.0	0.0	0.0	2.1	0.0	0.0	0.0	0.0
1,2	%		6.9	18.7	13.3	30.0	7.7	25.0	5.4	16.5	8.7	17.6	13.0	14.3
1-3	%		31.0	45.3	60.0	50.0	38.5	43.8	23.7	41.2	47.8	41.2	39.1	35.7

APPEARS THIS WAY
ON ORIGINAL

Appendix Table A.14. Study 376-404: Response Profiles in Lesion Counts: Completers

As noted in the section II.C, to control the Type I error associated with analyzing multiple visits for each lesion count or global assessment, these tests should be read as a closed family. That is, the results at week 9 are evaluated if and only if the test of difference between vehicle and Estrostep are statistically significant at week 10. Similarly the week 8 results are evaluated if and only if the test of difference between vehicle and Estrostep are statistically significant at week 9. If at any visit the differences are not statistically significant, the analysis stops and reports any observed statistical significance up to that time point.

Study 404

	VISIT																	
	03		04		05		06		07		08		09		10		eos	
	Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est
N	135	133	126	122	120	116	112	114	102	110	100	100	94	105	103	111	148	147
Inflammmatory																		
LS Mean	21.9	22.4	19.5	19.2	18.2	16.2	16.9	14.0	15.5	12.3	16.3	10.2	14.8	10.8	14.3	9.8	15.5	12.2
Std Err	0.8	0.8	1.0	1.0	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.8	0.8	0.9	0.9
p-value		.7072		.8346		.1206		.0300		.0116		.0000		.0023		.0001		.0081
Comedones																		
LS Mean	35.7	34.9	31.7	30.7	33.4	27.5	30.4	23.7	29.4	23.8	30.7	23.2	28.5	23.1	28.9	20.6	29.1	23.0
Std Err	1.2	1.1	1.4	1.4	1.6	1.6	1.3	1.3	1.4	1.4	1.5	1.5	1.6	1.5	1.5	1.5	1.3	1.3
p-value		.8350		.8001		.0131		.0010		.0088		.0008		.0203		.0002		.0018
Total Lesions																		
LS Mean	57.6	57.3	51.1	49.9	51.5	43.7	47.2	37.7	44.9	36.1	47.0	33.4	43.3	33.9	43.2	30.5	44.6	35.2
Std Err	1.5	1.4	1.8	1.8	1.9	2.0	1.8	1.8	1.9	1.9	2.0	2.0	2.0	1.9	1.9	1.9	1.8	1.8
p-value		.9561		.7672		.0075		.0003		.0018		.0000		.0011		.0000		.0004
% Change in Inflammatory Lesions																		
LS Mean	25.6	22.9	34.1	33.9	38.8	44.4	43.5	51.4	47.3	58.4	44.7	65.4	48.3	62.5	51.0	66.4	45.8	58.8
Std Err	3.0	2.9	3.3	3.3	3.1	3.2	3.2	3.2	3.1	3.0	3.1	3.1	3.4	3.1	2.8	2.7	3.0	2.9
p-value		.5269		.9894		.1994		.0942		.0110		.0000		.0026		.0001		.0022
% Change in Comedones																		
LS Mean	10.1	9.7	19.4	20.2	18.5	27.7	22.0	36.6	24.3	37.4	20.4	40.2	25.3	38.3	25.1	43.9	23.6	38.3
Std Err	3.4	3.3	3.9	3.9	3.9	4.0	3.6	3.6	4.4	4.3	4.4	4.4	4.6	4.3	4.6	4.4	3.7	3.6
p-value		.8387		.9142		.1130		.0052		.0366		.0022		.0418		.0034		.0046
% Change in Total Lesions																		
LS Mean	18.4	16.6	27.3	27.6	28.5	35.9	33.4	44.9	36.8	48.1	33.3	52.3	37.1	51.1	38.4	56.1	34.6	49.1
Std Err	2.3	2.3	2.7	2.7	2.8	2.8	2.6	2.6	3.0	2.9	3.0	3.0	3.2	3.0	2.9	2.8	2.6	2.6
p-value		.4632		.9617		.0821		.0030		.0089		.0000		.0018		.0000		.0001

APPEARS THIS WAY
ON ORIGINAL

Appendix Table A.15. Study 376-404: Response Profiles in Global Acne Assessment : Completers

To maintain overall error p-values are to be read from the right to the left, stopping at the first nonsignificant p-value.

	VISIT															
	03		04		05		06		07		08		09		10	
	Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est
1	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.
%	.	.	.	.	.	.	.	.	.	.	.	.	.	1.0	.	.
2	2	2	2	3	1	7	2	11	4	14	5	13	5	14	10	24
%	1.5	1.5	1.6	2.5	0.8	6.0	1.8	9.6	3.9	12.7	5.0	12.9	5.3	13.3	9.7	21.6
3	16	15	21	23	21	22	25	28	25	28	22	35	26	33	26	34
%	11.7	11.0	16.4	18.9	17.5	19.0	22.3	24.6	24.5	25.5	22.0	34.7	27.7	31.4	25.2	30.6
4	41	46	47	36	42	38	34	39	33	35	29	28	28	37	36	34
%	29.9	33.8	36.7	29.5	35.0	32.8	30.4	34.2	32.4	31.8	29.0	27.7	29.8	35.2	35.0	30.6
5	55	59	43	48	40	41	34	26	26	28	35	21	24	15	20	17
%	40.1	43.4	33.6	39.3	33.3	35.3	30.4	22.8	25.5	25.5	35.0	20.8	25.5	14.3	19.4	15.3
6	21	14	12	12	13	7	13	10	11	5	8	4	7	5	7	2
%	15.3	10.3	9.4	9.8	10.8	6.0	11.6	8.8	10.8	4.5	8.0	4.0	7.4	4.8	6.8	1.8
7	2	.	3	.	3	1	4	.	3	.	1	.	4	.	4	.
%	1.5	.	2.3	.	2.5	0.9	3.6	.	2.9	.	1.0	.	4.3	.	3.9	.
N	135	133	126	122	120	116	112	114	102	110	100	100	94	105	103	111
Global=1,2																
n	2	2	2	3	1	7	2	11	4	14	5	13	5	15	10	24
%	1.5	1.5	1.6	2.5	0.8	6.0	1.8	9.6	3.9	12.7	5.0	12.9	5.3	14.3	9.7	21.6
p-value	.937		.534		.012		.006		.011		.042		.018		.009	
Global=1,2,3																
n	17	17	22	26	22	29	27	39	29	42	27	48	31	48	36	58
%	12.6	12.8	17.5	21.3	18.3	25.0	24.1	34.2	28.4	38.2	27.0	47.5	33.0	45.7	35.0	52.3
p-value	.955		.361		.128		.059		.094		.001		.033		.004	
Global>3																
n	118	116	104	96	98	87	85	75	73	68	73	52	63	57	67	53

APPEARS THIS WAY
ON ORIGINAL

Appendix Table A.16. Study 376-404: Response Profiles: LOCF Subjects

To maintain overall error, p-values are to be read from the right to the left, stopping at the first nonsignificant p-value.

Study 404

	VISIT 03		VISIT 04		VISIT 05		VISIT 06		VISIT 07		VISIT 08		VISIT 09		VISIT 10	
	Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est
N	148	147	148	147	148	147	148	147	148	147	148	147	148	147	148	147
Inflammatory																
LS Mean	22.2	22.6	19.7	19.3	18.5	17.2	17.5	15.5	17.0	13.9	16.9	12.4	16.2	12.7	15.5	12.2
Std Err	0.8	0.8	0.9	0.9	0.9	0.9	0.9	0.9	1.0	0.9	0.9	0.9	0.9	0.9	0.9	0.9
p-value	.7151		.7817		.3393		.1438		.0236		.0006		.0071		.0081	
Comedones																
LS Mean	35.3	35.3	32.1	31.7	32.7	29.4	30.5	25.6	29.9	25.6	30.3	24.9	28.8	24.7	28.9	23.2
Std Err	1.1	1.1	1.2	1.2	1.4	1.4	1.2	1.2	1.2	1.2	1.2	1.2	1.3	1.2	1.3	1.3
p-value	.9978		.8273		.0912		.0042		.0137		.0022		.0206		.0018	
Total Lesions																
LS Mean	57.5	57.9	51.8	51.0	51.2	46.6	48.0	41.2	46.9	39.6	47.3	37.4	45.1	37.4	44.5	35.4
Std Err	1.4	1.3	1.7	1.6	1.9	1.8	1.7	1.7	1.8	1.7	1.8	1.7	1.8	1.7	1.8	1.8
p-value	.8362		.7464		.0791		.0043		.0034		.0001		.0022		.0004	
% Change in Inflammatory																
LS Mean	24.7	22.7	33.1	34.4	37.2	41.6	40.5	46.6	41.1	52.7	41.2	57.9	43.5	57.0	45.9	58.9
Std Err	2.8	2.7	2.9	2.9	3.1	3.0	3.1	3.1	3.1	3.0	3.1	3.0	3.1	3.0	3.0	2.9
p-value	.6185		.7525		.3061		.1627		.0083		.0002		.0021		.0022	
% Change in Comedones																
LS Mean	9.9	8.7	18.1	18.2	18.7	22.2	20.5	32.0	21.9	32.1	20.3	35.5	24.2	35.2	23.8	38.3
Std Err	3.1	3.0	3.5	3.4	3.6	3.5	3.2	3.1	3.5	3.4	3.5	3.4	3.5	3.4	3.6	3.5
p-value	.7839		.9883		.4901		.0093		.0391		.0023		.0232		.0046	
% Change in Total Lesions																
LS Mean	17.8	15.5	25.8	26.1	27.7	31.7	30.8	39.8	32.2	42.5	31.2	46.2	34.0	46.2	35.0	49.1
Std Err	2.1	2.1	2.4	2.4	2.7	2.6	2.4	2.4	2.6	2.5	2.6	2.5	2.6	2.6	2.6	2.5
p-value	.4416		.9345		.2902		.0084		.0049		.0001		.0010		.0001	

	VISIT 03		VISIT 04		VISIT 05		VISIT 06		VISIT 07		VISIT 08		VISIT 09		VISIT 10	
	Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est
N	135	133	139	136	139	136	139	136	139	136	139	136	139	136	139	136
Global=1,2																
n	2	2	2	3	2	7	3	11	5	15	6	15	7	17	11	25
%	1.5	1.5	1.4	2.2	1.4	5.1	2.2	8.1	3.6	11.0	4.3	11.0	5.0	12.5	7.9	18.4
p-value	.937		.563		.042		.016		.009		.024		.019		.006	
Global=1,2,3																
n	17	17	23	28	24	33	30	44	34	48	33	59	40	57	40	63
%	12.6	12.8	16.5	20.6	17.3	24.3	21.6	32.4	24.5	35.3	23.7	43.4	28.8	41.9	28.8	46.3
p-value	.955		.310		.094		.034		.032		.001		.019		.002	
Global>3																
n	118	116	116	108	115	103	109	92	105	88	106	77	99	79	99	73

Appendix Table A.17. Study 404: Subgroup Lesion Counts

Visit		10 (Completers)				eos (ITT)			
Age Group		19-		20+		19-		20+	
		Veh	Est	Veh	Est	Veh	Est	Veh	Est
	N	41	39	62	72	54	51	94	96
Inflammatory Lesions	Mean	17.6	11.3	11.5	8.6	17.9	12.9	14.1	12.1
	Std Error	2.2	1.4	1.1	0.8	1.7	1.3	1.0	1.3
Comedones	Mean	35.0	20.7	26.4	21.1	34.1	24.7	27.4	23.5
	Std Error	3.2	2.2	2.3	1.8	2.6	2.9	1.8	1.7
Total Lesions	Mean	52.6	32.0	37.9	29.7	52.1	37.5	41.5	35.6
	Std Error	4.6	3.0	3.0	2.3	3.7	3.9	2.4	2.5
% Change in Inflam. Lesions	Mean	44.1	61.3	57.9	71.0	43.3	56.5	46.9	59.2
	Std Error	6.8	4.7	3.6	2.7	5.4	4.3	3.6	4.0
% Change in Comedones	Mean	21.0	43.3	24.3	42.4	21.6	38.6	20.3	35.4
	Std Error	6.6	8.5	6.4	5.0	5.4	7.1	4.8	4.4
% Change in Total Lesions	Mean	33.1	53.3	41.6	57.4	32.6	48.1	33.9	47.9
	Std Error	4.7	5.0	4.1	2.6	4.0	4.5	3.5	3.1

Visit		10 (Completers)						eos (ITT)					
Race		White		Black		Other		White		Black		Other	
		Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est
	N	74	79	13	13	16	19	105	103	21	20	22	24
Inflammatory Lesions	Mean	13.2	10.1	11.4	7.3	19.4	8.6	14.7	12.8	15.3	12.7	19.3	10.1
	Std Error	1.2	0.9	2.3	2.6	4.2	1.6	1.0	1.2	2.2	2.8	3.1	1.5
Comedones	Mean	30.6	20.7	25.8	19.0	29.7	23.4	29.5	23.4	30.2	26.5	31.0	23.9
	Std Error	2.4	1.6	3.1	2.7	4.9	4.3	1.9	1.8	2.5	4.2	4.0	3.6
Total Lesions	Mean	43.8	30.8	37.2	26.3	49.1	32.0	44.3	36.3	45.5	39.2	50.3	34.0
	Std Error	3.2	2.1	4.5	4.6	8.0	5.6	2.5	2.6	3.9	6.2	6.2	4.6
% Change in Inflam. Lesions	Mean	54.7	64.9	54.7	73.2	39.8	74.4	49.2	56.4	39.6	57.1	34.2	67.0
	Std Error	3.9	3.1	9.0	8.1	11.8	3.7	3.3	3.9	9.6	8.1	9.2	4.4
% Change in Comedones	Mean	23.0	42.6	23.9	51.4	22.1	37.2	23.2	36.0	12.3	36.6	17.7	38.4
	Std Error	5.8	5.0	10.3	6.5	10.3	15.0	4.4	4.5	8.6	6.4	8.2	12.0
% Change in Total Lesions	Mean	39.5	54.5	38.9	62.5	31.3	57.4	36.5	46.9	25.9	46.4	26.1	53.9
	Std Error	3.7	3.0	8.2	4.6	8.7	6.5	3.0	3.2	7.5	6.3	7.2	5.6

APPEARS THIS WAY
ON ORIGINAL

Appendix Table A.18. Study 403: Subgroup Global Evaluations

Visit			10 (Completers)				eos (ITT)			
Age Group			19-		20+		19-		20+	
			Veh	Est	Veh	Est	Veh	Est	Veh	Est
2. Minimal	n		3	9	7	15	3	9	8	16
	%		7.3	23.1	11.3	20.8	5.7	18.4	9.2	17.4
3. Mild	n		7	7	19	27	7	11	24	33
	%		17.1	17.9	30.6	37.5	13.2	22.4	27.6	35.9
4. Mild to Moderate	n		15	13	21	21	21	15	29	25
	%		36.6	33.3	33.9	29.2	39.6	30.6	33.3	27.2
5. Moderate	n		9	8	11	9	14	12	17	15
	%		22.0	20.5	17.7	12.5	26.4	24.5	19.5	16.3
6. Marked	n		4	2	3	.	5	2	8	3
	%		9.8	5.1	4.8	.	9.4	4.1	9.2	3.3
7. Severe	n		3	.	1	.	3	.	1	.
	%		7.3	.	1.6	.	5.7	.	1.1	.
1,2	%		7.3	23.1	11.3	20.8	5.7	18.4	9.2	17.4
1-3	%		24.4	41.0	41.9	58.3	18.9	40.8	36.8	53.3

Visit			10 (Completers)						eos (ITT)					
Race			White		Black		Other		White		Black		Other	
			Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est	Veh	Est
2		n	7	18	1	4	2	2	8	19	1	4	2	2
	%		9.5	22.8	7.7	30.8	12.5	10.5	8.0	19.0	5.0	23.5	10.0	8.3
3		n	19	22	4	5	3	7	23	31	4	5	4	8
	%		25.7	27.8	30.8	38.5	18.8	36.8	23.0	31.0	20.0	29.4	20.0	33.3
4		n	28	25	4	3	4	6	37	28	7	4	6	8
	%		37.8	31.6	30.8	23.1	25.0	31.6	37.0	28.0	35.0	23.5	30.0	33.3
5		n	12	12	4	1	4	4	20	18	7	4	4	5
	%		16.2	15.2	30.8	7.7	25.0	21.1	20.0	18.0	35.0	23.5	20.0	20.8
6		n	5	2	.	.	2	.	9	4	1	.	3	1
	%		6.8	2.5	.	.	12.5	.	9.0	4.0	5.0	.	15.0	4.2
7		n	3	.	.	.	1	.	3	.	.	.	1	.
	%		4.1	.	.	.	6.3	.	3.0	.	.	.	5.0	.
1,2	%		9.5	22.8	7.7	30.8	12.5	10.5	8.0	19.0	5.0	23.5	10.0	8.3
1-3	%		35.1	50.6	38.5	69.2	31.3	47.4	31.0	50.0	25.0	52.9	30.0	41.7

APPEARS THIS WAY
ON ORIGINAL

2 page(s) have been removed because it contains trade secret and/or confidential information that is not disclosable.

Statistical Review and Evaluation
Addendum

NDA/ Drug Class:	21-276 / 6S
Name of Drug:	Estrostep® (norethindrone/acetate/ethynyl estradiol) Tablets
Applicant:	Parke-Davis Pharmaceutical Research
Indications:	Treatment of moderate acne vulgaris in females age 14 or greater who desire contraception.
Documents Reviewed:	Volumes in FDA Electronic Data Room, including SAS data sets, plus paper volumes 1.2, 1.33-155, submitted July 3, 2000.
Medical Officer:	Joseph Porres, M.D., Ph.D. (HFD-540)
Statistical Reviewer:	Steve Thomson (HFD-725)

Background:

Following a meeting between the clinical and statistical teams to discuss labeling, the medical team recommended that the label contain efficacy results for the combined pivotal studies 376-403 and 376-404. However, the Statistical Review, dated March 8, 2001 evaluated efficacy for each study separately. The purpose of this addendum to the Statistical Review is to combine and summarize the efficacy results for the two studies as per the clinical team request.

Table 1. Combined Efficacy Results for Pivotal Studies 376-403 and 376-[redacted]
Primary Efficacy Variables Mean Change in Number of Lesions at Six Months From Baseline*

[redacted]

APPEARS THIS WAY
ON ORIGINAL

Steve Thomson
Mathematical Statistician, Biometrics III

concur: **Mohamed Alosh, Ph.D.**
Team Leader, Biometrics III

cc:
Archival NDA: 21-276 Estrostep
HFD-540/Division File
HFD-540/Dr. Wilkin
HFD-540/Dr. Walker
HFD-540/Dr. Porres
HFD-540/Ms. Cintron
HFD-725/Dr. Huque
HFD-725/Dr. Alosh
HFD-725/Mr. Thomson
This review has 2 pages including this signature page..
Chron.